Biochemical investigations of TetR-family regulators that control antibiotic biosynthesis in *Streptomyces* bacteria

Alexander James Fullwood

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Life Sciences

School of Life Sciences, University of Warwick

Supervisors: Christophe Corre, Richard Napier and Vilmos Fülöp

June 2019

Table of Contents

List of Figures

List of Tables

Acknowledgements

Firstly, I would like to express my most sincere gratitude to my supervisor Dr Christophe Corre for his patient, thoughtful, knowledgeable and empathetic guidance and support in keeping me on the right track, as well as his patience in dealing with my inexperience in this area of expertise. I would also like to sincere thanks to Prof Richard Napier for his expert guidance and *in vitro* protein biochemistry and Surface Plasmon Resonance, and both Prof Vilmos Fülöp and Prof Alex Cameron for their advice regarding protein crystallography and the use of the LCP Mosquito equipment.

From the Corre group I would like to give thanks to Vincent Poon and Kathryn Styles for their guidance in the day-to-day running of lab C126a, as well as additional thanks to Vincent for providing both necessary research materials and bouncing ideas off based on bioinformatics and experimental work in *Streptomyces avermitilis*. I wish to also express my gratitude towards Dr John Sidda, Dr Miriam Rodríguez García, Jehad Aldali, Dr Fabrizio Alberti and other members of the Corre group both past and present who made my time at Warwick all that much more enjoyable.

From the Napier group I would like to especially thank Mussa Quareshy and Justyna Prusińska for their guidance in the day-to-day running of protein purification protocols and equipment, as well as working of the Biacore T-200 equipment. From the Challis group I would like to thank Shanshan Zhou for her advice and feedback of EMSA experimental protocols, as well as the provision of materials necessary for some of the work in my thesis. And finally, I'd like to give thanks to my family for their continual love and support during my studies.

Declarations

This thesis was submitted to the University of Warwick in support of the authors application for the degree of Doctor of Philosophy. Original research, analysis and composition of the thesis was carried out by the author except for the work stated below which supports the material contained herein:

- Crystal structures of *Streptomyces coelicolor* A3(2) MmfR in its ligand-free and ligand-bound forms used for *in silico* studies as described in Chapter 5 and Chapter 6 were solved by Dr Dean Rea (manuscript in progress).
- Microbial cultures of *Streptomyces avermitilis* and *Streptomyces sclerotialus* used for polymerase chain reaction (PCR) experiments were cultivated by Dr Vincent Poon and Jacqueline Gill respectively.
- pET151 constructs encoding SgnR, MmfR and MmfR mutants were provided by Dr Shanshan Zhou.
- Organic synthesis of signalling molecules used in electrophoretic mobility shift assays (EMSAs) and surface plasmon resonance (SPR) experiments as described in Chapter 3 and Chapter 6 were synthesised by Dr Zhou.
- Similarly, methylenomycin A (MmA) was purified from *Streptomyces albus* J1074 strain W303 by Dr Gideon Idowu.
- Bioinformatics performed in Chapter 5 were performed in collaboration with Dr Poon.

Parts of this thesis have been published by the author:

Bowyer, J.E., de los Santos, ELC., Styles, K.M., Fullwood, A.J., Corre, C., Bates, D.G., Modeling the architecture of the regulatory system controlling methylenomycin production in *Streptomyces coelicolor*. *Journal of Biological Enginee*ring. **2017**, *11:12*.

Abstract

Many gene clusters encoding secondary metabolites biosynthesis in *Streptomyces* bacteria are controlled by complex regulatory networks, producing silent phenotypes when *Streptomyces* are grown under standard laboratory conditions. 2-alkyl-4-hydroxymethylfuran-3-carboxylic acids (AHFCAs) induce secondary metabolite production in *Streptomyces* spp. by dissociating ArpA-subfamily transcriptional repressors from pseudo-palindromic autoregulator responsive elements (AREs).

In *Streptomyces coelicolor* A3(2) the AHFCA receptor MmfR and the pseudo-AHFCA receptor MmyR regulate the methylenomycin (*mmy*) gene cluster in tandem by controlling the expression of the *mmy* transcriptional activator *mmyB* as well as a 5-gene AHFCA regulatory cassette encoding *mmfR*, *mmyR* and the three AHFCA biosynthetic enzymes *mmfLHP*. Orthologous regulatory cassettes can be found within gene clusters of novel antibacterial interest across many *Streptomyces* spp., including *Streptomyces venezuelae*, *Streptomyces avermitilis* and *Streptomyces sclerotialus*.

The aim of this project was to characterise *in vitro* both the DNA-binding mechanisms and AHFCA structure-activity relationships (SAR) of MmfR/MmyR orthologues found in *S. avermitilis* (AvaL1/AvaL2) and *S. sclerotialus* (SclM1/SclM4). All orthologues excluding SclM4 were solubilised *in vitro*, and crystallisation trials yielded crystals of AvaL1, which remain to be diffracted. Gel shift assays demonstrated AvaL1/SclM1 binding to three/four AREs within their respective gene clusters, matching the 24-bp MmfR-type ARE consensus sequence 5'-AAnATACCTTCG|CGAAGGTATnTT-3'.

Surface plasmon resonance (SPR) established that AvaL1 bound to the three *S. avermitilis* AREs as a pair of homodimers with a mean K_D of 123.7 nM ± 11.2 nM at 25 °C. Kinetic analysis inferred a two state conformation-dependent DNA-binding mechanism likely associated with the stoichiometry of bound dimers. AvaL1 binding to DNA was more significantly impaired in the presence of endogenous *S. avermitilis* AHFCAs (mean IC_{50} = 0.48 μM ± 0.06 μM), rationalised *in silico* by enlargement of the AvaL1 ligand-binding pocket. The pseudo-AHFCA receptor AvaL2 was also shown not bind to the same AREs as AvaL1, and its effector remains unknown.

Abbreviations, Acronyms and Initialisms

ACT/*act* – Actinorhodin antibiotic/gene cluster
ADV – AutoDock Vina
AHFCA – 2-alkyl-4-hydroxymethylfuran-3-carboxylic acid
AfsA – A-factor biosynthesis protein A
AMP – Ampicillin
ANOVA – Analysis of Variance
antiSMASH – Antibiotics and Secondary Metabolite Analysis Shell
ARE – Autoregulator-responsive element
ArpA – A-factor receptor protein A
***ArpA*$_{CS}$** – ArpA consensus sequence
AUC – Analytical Ultracentrifugation
avaBC – Contraction of the *avaB-avaC* operon
AVE/*ave* – Avermectin antibiotic/gene cluster
BLAST – Basic Local Alignment Search Tool
BM – Best Match
BprA – γ-butyrolactone putative reductase A
CAPS – N-cyclohexyl-3-aminopropanesulfonic acid
CASTing – Cyclic Amplification and Selection of Targets
CD – Circular Dichroism
CDA/*cda* – Calcium-dependant antibiotic/gene cluster
ChIP – Chromatin Immunoprecipitation
CO – Codon-optimised
CPK/*cpk* – Cryptic type I polyketide synthase antibiotic/gene cluster
CprA – *S. coelicolor* A3(2) pleiotropic regulator A
CprB – *S. coelicolor* A3(2) pleiotropic regulator B
CSI – Cognate Site Identifier
CryoEM – Cryogenic Electron Microscopy
CS – Consensus Sequence
DBD – DNA-Binding domain
DMSO – Dimethyl sulfoxide
DNA – Deoxyribonucleic acid
dNTPs – Deoxyribonucleoside triphosphate
DTT – Dithiothreitol
EDTA – Ethylenediaminetetraacetic acid
EMSA – Electrophoretic Mobility Shift Assay
FA/FP – Fluorescence Anisotropy/Polarisation
Fc – Flow Cell
FIMO – Find Individual Motif Occurrences
GBL – Gamma(γ)-butyrolactone
GBN/*gbn* – Gaburedin antibiotic/gene cluster
HADDOCK – High Ambiguity Driven protein-protein DOCKing
HEPES – 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HFCA – 4-hydroxymethylfuran-3-carboxylic acid
HGT – Horizontal Gene Transfer
HTH – Helix-turn-helix
IMAC – Immobilised Metal Affinity Chromatography

IPTG – Isopropyl β-D-1-thiogalactopyranoside
ITC – Isothermal Titration Calorimetry
IUPAC – International Union of Pure and Applied Chemistry
JCSG – Joint Centre for Structural Genomics
jad – Jadomycin gene cluster
JdA – Jadomycin A
JdB – Jadomycin B
LB – Lysogeny broth
LBA – Lysogeny broth on agar plate
LBD – Ligand-Binding Domain
LPC – Lipidic Cubic Phase
MAD – Multi-wavelength Anomalous Dispersion
MALS – Multi-Angle Light Scattering
MAST – Motif Alignment and Search Tool
MBP – Maltose Binding Protein
MEME – Multiple Em for Elucidation of Motifs
MES – 2-(N-morpholino)ethanesulfonic acid
MIDAS – Modern Intelligent Dynamic Alternative Screen
MITOMI – Mechanically Induced Trapping of Molecular Interactions
ML – Maximum Likelihood
MmA – Methylenomycin A
MmB – Methylenomycin B
MmC – Methylenomycin C
MME – Monomethyl ether
MMF – Methylenomycin Furan
MmfR – MMF response regulator
MmfL – MMF GBL biosynthesis protein
mmfLHP – Contraction of the *mmfL-mmfH-mmfP* operon
MmfH – MMF hydrolase
MmfP – MMF phosphatase
mmy – Methylenomycin gene cluster
MmyR – Methylenomycin gene cluster regulator
MOPS – 3-(N-morpholino)propanesulfonic acid
MPD – 2-Methyl-2,4-pentanediol
MR – Molecular Replacement
mRNA – Messenger RNA
MUSCLE – Multiple Sequence Comparison by Log Expectation
MWCO – Molecular Weight Cut-Off
NTA – Nitrilotriacetic acid
OPB – Operator for CprB
ORF – Open Reading Frame
PACT – pH, anion, cation crystallization trial
PBM – Protein Binding Microarray
PCR – Polymerase Chain Reaction
PDB – Protein Databank *file*
PEG – Polyethylene glycol
PHYLIP – PHYLogeny Inference Package
PyMOL – Python™ Molecule Visualisation System
qPCR – Quantitative (real-time) PCR
RCF – Relative Centrifugal Force
RED/*red* – Undecylprodigiosin antibiotic/gene cluster
RMSD – Root-Mean-Square Deviation
RNA – Ribonucleic acid
RNAP – RNA Polymerase
RU – Response Units
SA – Streptavidin

SAR – Structure-Activity Relationship
SEC – Size-Exclusion Chromatography
SELEX – Systematic Evolution of Ligands by Exponential Enrichment
SOC – Super Optimal broth with Catabolite repression
SCB – *Streptomyces coelicolor* A3(2) butanolide
ScbR – SCB response regulator
ScbR2 – Pseudo-SCB response regulator
SCL/*scl* – Scleric acid antibiotic/gene cluster
sclM2L3 – Contraction of the *sclM2-sclL-sclM3* operon
sclN-E – Contraction of the *sclN*-through-*sclE* operon
sclQ1-4P – Contraction of the *sclQ1-sclQ2-sclQ3-sclQ4-sclP* operon
SCP – *Streptomyces coelicolor* A3(2) plasmid
SwD – SwissDock
SDS-PAGE – Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEM – Standard Error of the Mean
SPR – Surface Plasmon Resonance
TAR – Transformation-Associated Recombination
TetR – Tetracycline resistance regulator
TEV – Tobacco Etch Virus
TFR – TetR-family transcriptional repressor
TOPO – Topoisomerase
tRNA – Transfer RNA
UPGMA –Unweighted Pair Group Method with arithmetic Mean
UV – Ultraviolet
WISB - Warwick Integrative Synthetic Biology

1 Introduction

1.1 *Streptomyces*, the antibacterial workhorse

1.1.1 Microbiology of *Streptomyces* bacteria

Streptomyces are aerobic, filamentous Gram-positive bacteria belonging to the phylum *Actinobacteria* within the order of *Actinomycetales* (actinomycetes), which vary greatly in their morphology, physiology and biochemistry. These bacteria are so named based on their morphological similarity to fungi (-mycete, from Greek for mushroom or fungi), ensuring long-term survival by producing mycelia and aerial hyphae *via* dispersion of bacterial spores (Figure 1-1). *Streptomyces* are found abundantly in soil microbiomes (up to 90 % of all actinomycetes) decomposing organic matter as source of energy (saprophytic), forming a major fraction of the humus layer. Although soil *Streptomyces* are the most studied, marine *Streptomyces species*, such as *Streptomyces maritimus* can be found in seawater, sediment, fishes, molluscs, sponges, seaweeds and mangroves[(1)].

Few *Streptomyces* spp. are pathogenic; some species are phytotoxic, causing potato crop diseases such as such as common scab (*Streptomyces scabies*, *Streptomyces turgidiscabies*, *Streptomyces stelliscabiei*), netted scab (*Streptomyces europaeiscabiei*, *Streptomyces reticuliscabiei*), acid scab (*Streptomyces acidiscabies*), Russet scab (*Streptomyces cheloniumii*), and sweet potato soil rot (*Streptomyces ipomoeae*)[(2)]. *Streptomyces* spp. are also reported to cause pod wart in peanuts[(3)]. Potato diseases caused by these *Streptomyces* spp. share a common origin , primarily through production of phytotoxic diketopiperazine compounds such as thaxtomin[(4)], the biosynthetic genes of which are encoded on a mobile pathogenicity island likely originating from *S. scabies*[(5)]. Other compounds that contribute to potato pathogenesis are concanamycins, FD-891 and borrelidin[(2)]. *Streptomyces avermitilis* produces anti-parasitic avermectin (AVE)-class compounds which are toxic to freshwater organisms and soil invertebrates[(6-8)], with a recently observed case of human nystagmus recorded by acute abamectin poisoning[(9)]. Opportunistic invasive *Streptomyces* spp. which have been found in humans include *Streptomyces coelicolor*, *Streptomyces pelletieri*, *Streptomyces griseus*, *Streptomyces bikiniensis*, *Streptomyces thermovulgarus*, *S. maritimus*, *Streptomyces olivaceus*, *Streptomyces albus*, *Streptomyces nobilis*, *Streptomyces parvulus*, *Streptomyces humidus*, *Streptomyces pilosus* and *Streptomyces indigoferus*[(10)]. However, *Streptomyces somaliensis* and *Streptomyces sudanensis* are among the few known human pathogenic species, contributing directly to actinomycosis[(11, 12)].

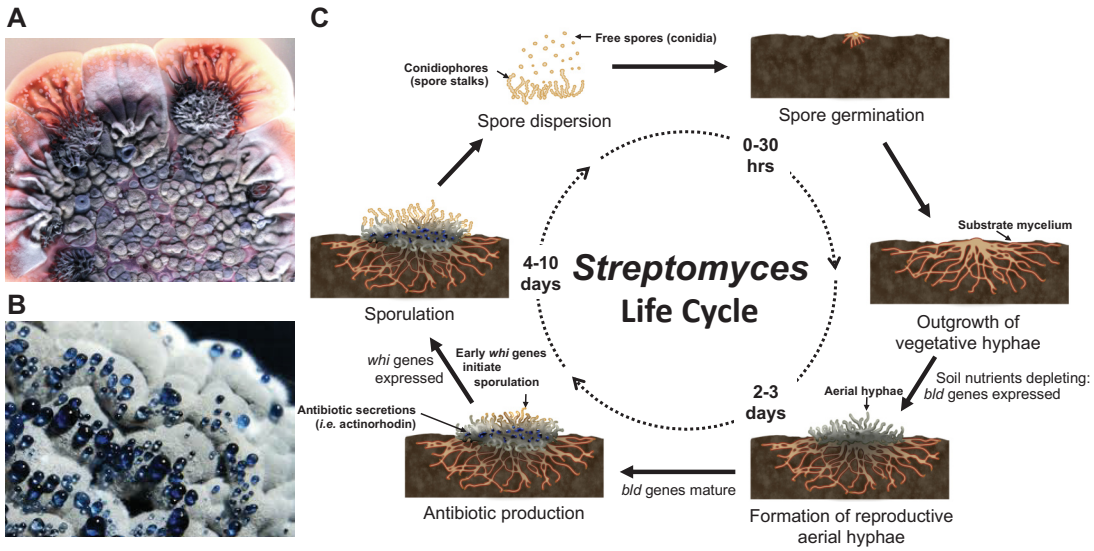


Figure 1-1 – *Streptomyces* morphology and life cycle. (**A**) The model *Streptomyces* organism *S. coelicolor* A3(2); image from "Exploring The Invisible" website[13]. (**B**) Sporulating *S. coelicolor* A3(2) colony producing the blue-pigmented antibiotic actinorhodin (ACT); image from Zacharia VM and Traxler MF, *Bacteria: Exploring new horizons*[14]. (**C**) Life cycle of *Streptomyces* bacteria.

1.1.2 The significance of *Streptomyces* in modern medicine

Streptomyces spp. are exploited in medicine for the wide range of effective and often complex secondary metabolites they produce. These include antibacterial, antiparasitic, antineoplastic (anticancer), antifungal, herbicidal and immunosuppressant compounds (Figure 1-2). These secondary metabolites are by definition not essential for bacterial growth, but instead allow *Streptomyces* bacteria to fulfil their ecological niche in complex soil/marine microbiomes as organic decomposers. This includes both the mediation of microbe–microbe interactions (*e.g.,* protection from antagonistic microbes, killing off competing microbes for limited resources, communication between mutualistic microbes, *etc.*)[15-18], as well as mediating symbiosis between *Streptomyces* bacteria and their hosts[19], which include plants[20-23], animals (*e.g.*, invertebrates and marine sponges)[24-28] and fungi/algae[29, 30], which can exploit the secondary metabolites produced by *Streptomyces* as protection against infection.

Streptomyces spp. have a long and complex history with antibacterial research and medicine. Approximately 70 % of clinically viable antimicrobial compounds originate from actinobacteria, of which *Streptomyces* spp. contributes the largest number[31]. Over half of these antibiotics were discovered in what is now referred to as the golden age of antibiotics, covering the period between 1940 and 1970, with a peak in the number of antibacterial discovered in the mid-to-late 1950s[31]. The first clinically significant *Streptomyces* antibiotic was streptomycin isolated from *S. griseus*. Streptomycin was discovered in 1943 by Selman and Waksman[32], for which Waksman would later receive the 1952 Nobel Prize in Physiology or Medicine. Streptomycin, the first-in-class aminoglycoside, is still used today in the

treatment of tuberculosis infections. However, the first secondary metabolite isolated from a *Streptomyces* sp. was the antineoplastic drug dactinomycin, isolated by Waksman and Woodruff from *Streptomyces antibioticus* in 1940[33]. Other *Streptomyces* antibiotics discovered in this era include the *Streptomyces* cephalosporins, chloramphenicol, neomycin, tetracycline, nystatin, viomycin, virginiamycin, lincomycin, cycloserine, novobiocin, rifamycin and kanamycin[31].

The discovery rate of antibiotics has decreased since the mid-1950s, leaving us in a "discovery void" since the late 1980s. Between 2010 and 2015 only eight new antibiotics have been approved by the US Food and Drug Administration[34], of which seven belong to existing drug classes, seven were more expensive than existing drug trial comparators, and only one, fidaxomicin, was superior over current antibiotics. With the advent of genome sequencing there is a new well of untapped potential for the discovery of novel compounds, but the difficulty of culturing bacteria in labs, of which 99 % of all bacteria are estimated to be unviable, greatly stifles progress in this field. In addition, metabolites produced by previously uncharacterised secondary metabolite pathways may result in the insolation of a compound which has already been previously characterised. This is exacerbated by the rapid proliferation of multiple-drug resistant strains of bacteria. However, all hope is not lost, as recently a novel compound, teixobactin, was isolated from *Eleftheria terrae* using modern technologies such as iChip to culture previously unculturable bacteria in soil[35].

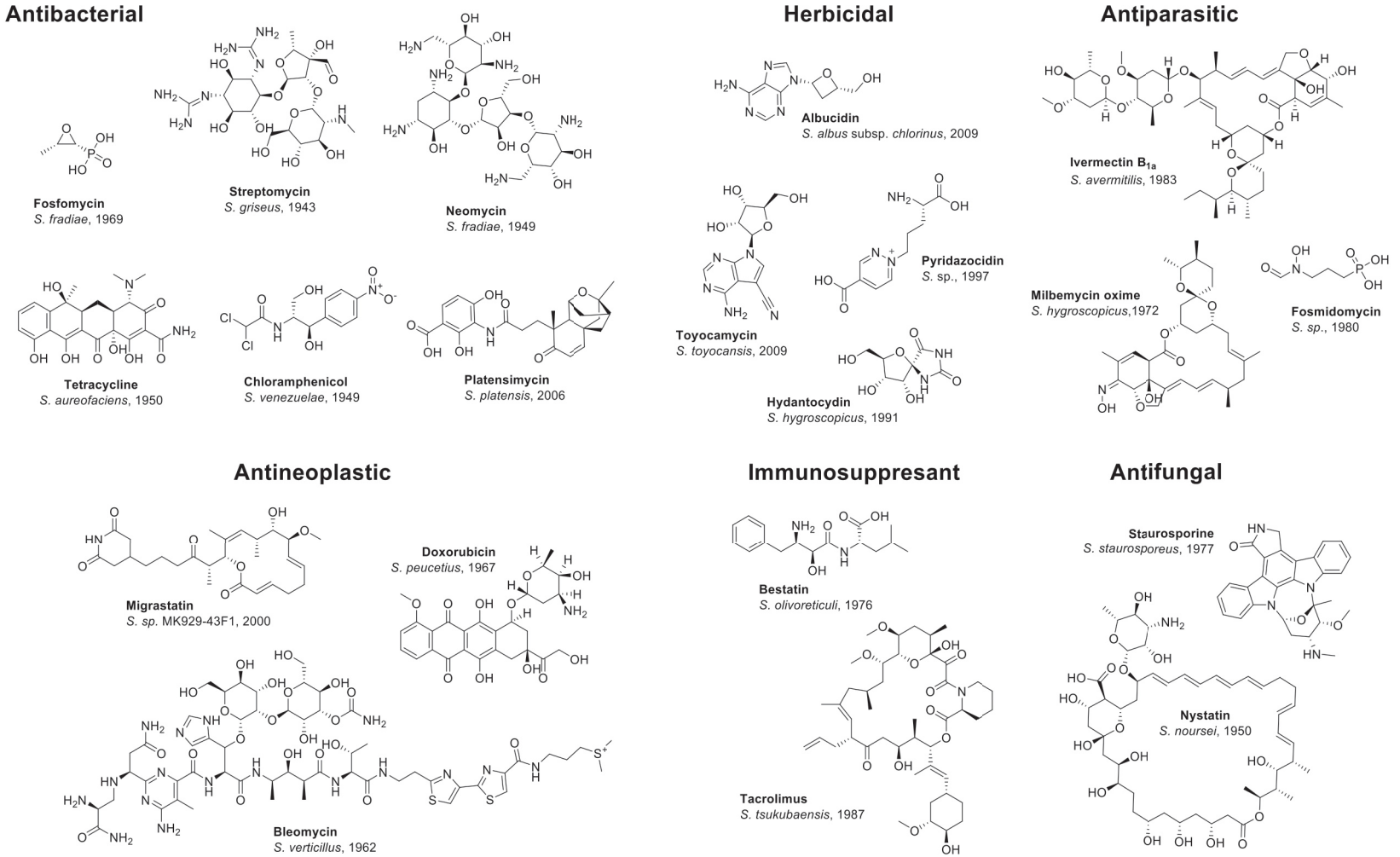


Figure 1-2 – Examples of medically and industrially significant chemical compounds produced by *Streptomyces* bacteria since the discovery of dactinomycin in 1940.

1.1.3 *Streptomyces coelicolor* A3(2), the model Streptomyces

S. coelicolor was first discovered growing on a potato by Müller in 1908 and originally classified as *Streptothrix coelicolor,* based on the production of a blue pigment which would later be identified as actinorhodin (ACT)[36]. A similar blue and red-pigment producing organism was discovered by Waksman in 1919, originally named *Actinomyces violaceus-ruber*. These later became *S. coelicolor* Müller and *S. violaceoruber* Waksman and Curtis, which at the time were thought to be synonymous with one another, but later shown to be two distinct organisms[37]. In modern nomenclature these strains are now *S. coelicolor* Müller and *S. coelicolor* A3(2). *S. coelicolor* today is frequently synonymous with *S. coelicolor* A3(2)), even though *S. coelicolor* A3(2) is now taxonomically a member of the species *S. violaceoruber*, while *S. coelicolor* Müller is more closely related to *S. griseus*.

In 2002 the genome of *S. coelicolor* M145 (*S. coelicolor* A3(2) lacking plasmids SCP1 and SCP2) was the first *Streptomyces* genome to be fully sequenced, and at the time comprised the largest sequenced bacterial genome[38]. Currently the largest reported bacterial genome is that of *Sorangium cellulosum* strain So0157-2 with has a genome size of 14.78-Mbp[39], and the smallest being Candidatus *Nasuia deltocephalinicola* with a genome size of only 112-kbp[40]. The 72.1 % GC linear genome of S. coelicolor A3(2) has a length of 8.67-Mbp and encodes 7,825 putative genes. The sequence of the 31-kbp plasmid SCP2 was published shortly after in 2003[41], and the 356-kbp plasmid SCP1 sequence in 2004[42]. The SCP2 plasmid is also among the earliest known examples of both antibiotic biosynthesis and resistance genes being encoded within a plasmid[43]. *S. avermitilis,* discovered in Japan in 1979[44], had in 2001 been sequenced with a genomic coverage of 99 %[45], with the full 9.03-Mbp genome being made available a year after *S. coelicolor* A3(2) in 2003[46].

1.1.4 Regulation of cryptic biosynthetic gene clusters

Even if a bacterium grown in a laboratory environment is among the few that can be cultured in laboratory conditions, this does not guarantee the production and identification of new metabolite products. Since the sequencing of the *S. coelicolor* A3(2) genome in 2002, using bioinformatics-based methods such as genome mining to identify and study potential bioactive metabolites has become increasingly prevalent, particularly with *Streptomyces* and other soil bacteria[47]. Genes responsible for producing both simple and complex secondary metabolites are frequently clustered together in bacterial genome; multiple hypotheses describe why this occurs. The most controversial hypothesis is the “selfish operon” model put forward by Lawrence and Roth in 1996, which states that genes cluster together as an evolutionary trait

optimising for the propagation of functionally coupled genes *via* horizontal gene transfer (HGT) between mutualistic bacteria[48]. This model is healthily disputed, and there is not yet a consensus amongst microbiologists. Most alternative hypotheses revolve around the assumption that clustering of genes results in an improvement of bacterial fitness; hypotheses include reduction of metabolic strain through minimisation of protein intracellular diffusion, and improved coregulatory efficiency by minimising the necessary regulatory information[49].

Many of these gene clusters are under the control of transcriptional regulators; these regulators may either directly and/or indirectly control gene expression, may be situated within the gene cluster itself, and may also regulate other biological pathways. It is not uncommon to find gene cluster-specific regulators which compete and/or cooperate with pleiotropic regulators that tie directly into nutrient or growth-phase dependent pathways[50]. Biosynthetic pathways can also be stimulated by the presence of an inducer, which include both antibiotics and signalling molecules of both endogenous or exogenous origins[51]. Another frequent prerequisite condition for antibiotic production is dependent on the state of the *Streptomyces* bacterium within its life cycle. Antibiotic production is generally activated during the transition between vegetative growth and production of sporulating aerial hyphae, the energy of which is sustained by metabolism of the substrate mycelium[52, 53]. Altogether, this means that until the prerequisite regulatory conditions are met, the compounds encoded within these gene clusters are not produced. These gene clusters are referred to as "silent" gene clusters, and much of the current difficulties in discovery of novel products is finding the keys that unlock these gene clusters. When the product of a gene cluster is uncharacterised or unknown, it is frequently referred to as a "cryptic" gene cluster.

The antibiotic gene clusters in *S. coelicolor* A3(2) which have been the most studied are the actinorhodin (*act*) gene cluster[54, 55], the undecylprodigiosin (*red*) gene cluster[55, 56], the coelimycin gene cluster (*cpk,* formerly *kas*)[57, 58], the SCP1 plasmid encoded methylenomycin (*mmy*) gene cluster[43] and the calcium-dependent antibiotic (*cda*) gene cluster[59] (Figure 1-3). Altogether there are ~30 biosynthetic gene clusters for secondary metabolites which have so far been characterised, but recent bioinformatics-based genome mining tools such as antiSMASH[60] predict there may be as many as 93 gene clusters in *S. coelicolor* A3(2).

The *act* gene cluster is regulated by multiple biological processes[50]. Regulators of the *act* gene cluster include DraR/AfsQ1 which bind to the *actII-3/actII-4* region, (repressive in a nitrogen-depleted environment), AdpA (repressive at high ADP/ATP concentrations), LexA (inhibitory when DNA is not damaged) and ROK7B7 (repressive in low xylose concentration environments)[50]. A rare leucine codon TTA is present in the coding frame of *actII-4* and other

genes, which are transcribed only in the presence of BldA, a regulatory tRNA which accumulates only during morphological differentiation (primarily the formation of aerial hyphae and sporulation)[61, 62]. Therefore, ACT is only produced at this transitional stage of the life cycle. In addition to these regulators, there are several other regulators of the *act* gene cluster that belong to the TetR-family of transcriptional regulators, such as the gene cluster-situated ActR (ActII-1), which regulates expression of the ACT exporter ActA (ActII-2), and is itself regulated by both ACT and precursor intermediates[61, 63].

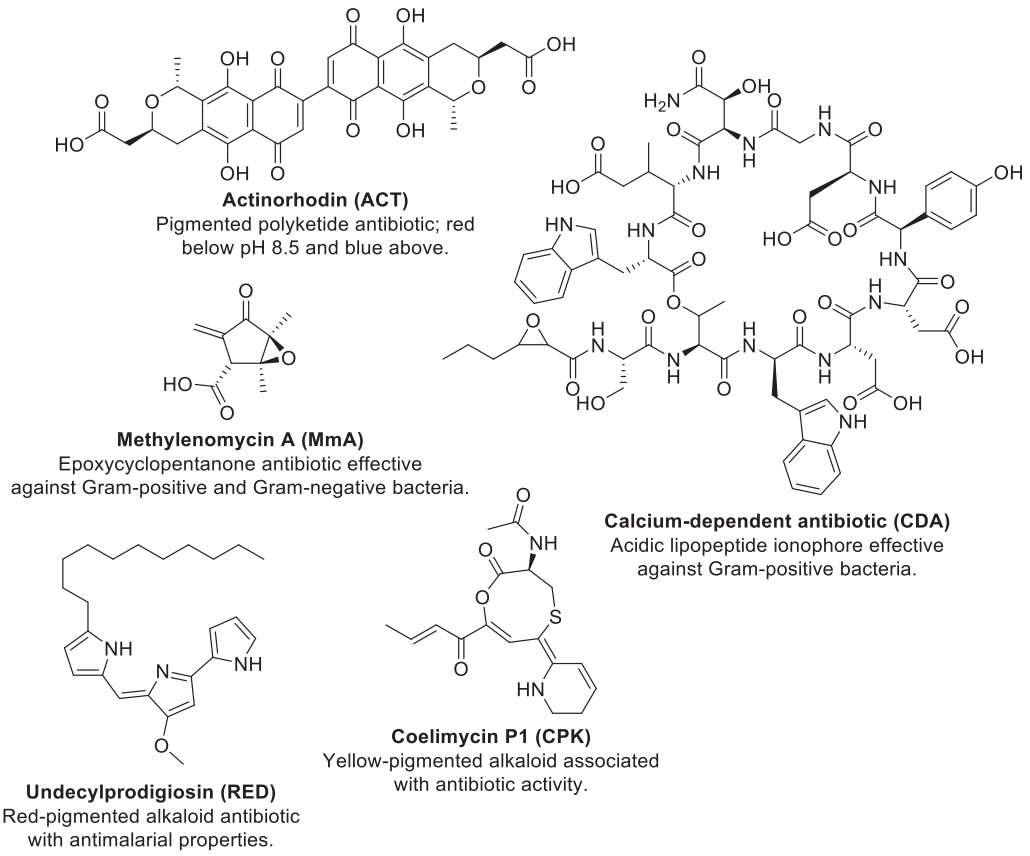


Figure 1-3 – Notable antibacterial secondary metabolites produced by *S. coelicolor* A3(2).

1.2 TetR-family transcriptional repressors (TFRs) regulate *Streptomyces* antibiotic biosynthesis

1.2.1 Molecular roles of TetR-family transcriptional repressors

The TetR-family transcriptional repressors (TFRs) are one-component signal transduction systems primarily acting as transcriptional repressors that regulate the expression of both their own gene and that of other target genes[64]. TFRs are homodimeric proteins, with each monomer consisting of an N-terminal DNA-binding domain (DBD) and a C-terminal ligand-binding domain (LBD). TFRs bind to pseudo-palindromic operators either downstream or overlapping promoter sites of target open reading frames (ORFs) as either a single dimer pair or as a pair of dimers, preventing transcription of the target genes by RNA polymerase (RNAP; Figure 1-4). TFRs are allosterically regulated by an effector; the effector, binds to the TFR LBD and induces a conformational change in the TFR dimer that releases the protein from DNA, enabling transcription of the target gene. TFRs are grouped into three types: type I (where TFRs genes are divergent from their target genes), type II (where TFR genes are encoded in the same operon as their target gene), and type III (neither type I or type II)[64]. Type I and Type II TFRs implicitly regulate their own expression (but not always). This typing convention does not necessarily apply to pleiotropic TFR regulators, as multiple TFR configurations are often observed.

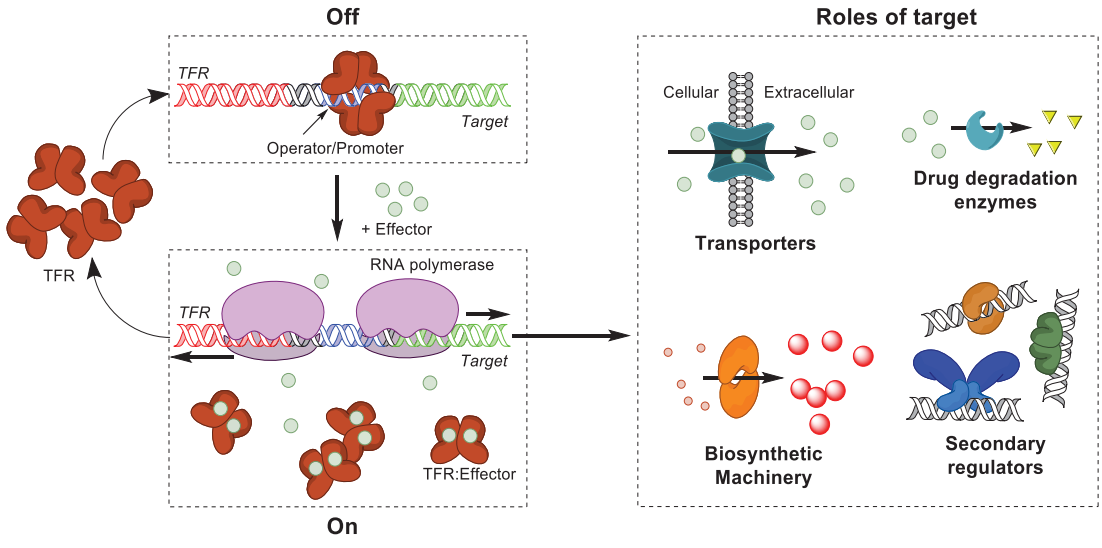


Figure 1-4 – Mode of action for a type I TetR-family transcriptional regulator (TFR). In this example homodimeric TFR binds at a single operator repressing the divergent expression of its own gene and that of the target gene (**off**). As effector concentration builds in the cytoplasm, binding of the effector to the TFR releases it from DNA, allowing expression of the target gene (**on**). As the level of intracellular TFR increases, the concentration of free effector is depleted and the expression of the TFR and its target is once again repressed (**off**).

The namesake protein is a tetracycline resistance regulator encoded in *Escherichia coli*, which controls the expression of the tetracycline efflux pump TetA by binding to the *tetO* operator in the absence of tetracycline. While TFRs are most frequently involved in drug resistance mechanisms by regulation of transporters and enzymes which deactivate antibacterial compounds, TFRs also play roles in pathways involved in co-factor metabolism, lipid metabolism, amino acid metabolism, nitrogen metabolism, carbon metabolism, lipid homeostasis, cytokinesis, heat-shock response, quorum sensing and secondary metabolite biosynthesis[64].

TFRs are abundant in *Streptomyces* spp.; in *S. coelicolor* A3(2) alone there are 151 TFRs encoded within the genome, accounting for 15.8 % of the 965 putative encoded regulatory proteins[38, 65]. As the model organism, the function of a number of these TFRs have already been investigated. The *act*, *cpk* and *mmy* gene clusters all contain TFRs with different functions; ActR regulates the *act* gene cluster as previously described, ScbR and ScbR2 control coelimycin biosynthesis by repressing *cpk* biosynthetic genes[66], and MmfR and MmyR repress *mmy* biosynthetic genes in a manner analogous to ScbR and ScbR2 (discussed in Chapter 1.3.1)[67].

TFRs can also function as pleiotropic regulators; regulating multiple targets to coordinate multiple pathways. The positive or negative activity of pleiotropic regulators against target pathways may be either through direct inhibition of target genes or through indirect inhibition by repression of secondary regulators which then either activate or repress antibiotic biosynthetic gene clusters. In *S. coelicolor* A3(2) the global pleiotropic regulators CprA and CprB regulate networks involved in secondary metabolite biosynthesis and morphological differentiation; CprA acts as a positive regulator of the *act*/*red* gene clusters and promotes sporulation[68], while CprB is a negative regulator of the *act*/*cpk* gene clusters and delays sporulation[68, 69]. *SCO3201* when overexpressed represses *cda*/*red*/*act* genes by indirect repression of other regulators[70]. AtrA binds directly upstream of the *act* transcriptional activator *actII-4* and inhibits ACT production[71]. XdhR directly binds to the promoters of *actR* and *actII-4* to also repress ACT production[72]. SCO1702 regulates either directly or indirectly the promoter regions of *actII-4*, *redD-Z* and *cdaR* to inhibit ACT, RED and CDA biosynthesis[73]. RrdA downregulates *redD* expression and upregulates *act* gene expression[74]. It is likely that other TFRs also play roles as pleiotropic regulators of these processes.

While some pleiotropic TFRs, such as CprA and CprB function specifically as pleiotropic regulators, other TFRs that primarily regulate a response to a given factor may also regulate other processes in response to that factor. For example, in the presence of guanosine

pentaphosphate ((p)ppGpp; an alarmone which accumulates during amino acid starvation) XdhR is released from DNA. In addition to allowing the expression of *actR* and *actII-4*, the *SCO1134-SCO1131 genes* neighbouring *xdhR*, are also expressed, which are involved in purine scavenging[75]. One possible interpretation is that *S. coelic*olor A3(2) produces actinorhodin under amino acid starved conditions to kill off other microorganisms competing for the same environmental resources.

In summary, TFRs often play a large role in the regulation of these complex systems. Because so many TFRs are encoded in *Streptomyces* spp. genomes, there may be many more cross-pathway regulators that remain to be identified.

1.2.2 ArpA-subfamily repressors coordinate antibiotic production *via* quorum sensing

The ArpA-subfamily repressors, named after the streptomycin biosynthesis regulator in *S. griseus*, are autoregulatory TFRs with two specialised functions[65]. Firstly, they frequently regulate genes involved in secondary metabolite (primarily antibiotic) biosynthesis and are often localised within the gene clusters they regulate. For gene clusters contain one or more ArpA regulators they are often the only common element conserved between *Streptomyces* spp., with the genes encoding for the metabolite biosynthesis often have little similarity with those of other gene clusters. Therefore, many ArpA-regulated gene clusters produce structurally different compounds. ArpA-subfamily repressors are also frequent pleiotropic regulators of other secondary pathways, which include antibiotic biosynthesis, morphological differentiation and stress response pathways[68, 76-79].

Secondly, ArpA-subfamily repressors recognise a class of small hydrophobic signalling hormones termed collectively in the literature as gamma(γ)-butyrolactones (GBLs), although strictly speaking this name does not apply to all ArpA signalling molecules as these compound structures can vary greatly (Figure 1-5). The biosynthetic genes for these effectors which are involved in colonial cell-to-cell quorum sensing *in situ*, are also encoded within ~60 % of *Streptomyces* spp. genomes[80], typically situated near the gene(s) of GBL-receptive ArpA-subfamily repressors. ArpA regulators frequently represses expression of these GBL biosynthetic genes, forming a self-regulating genetic unit referred to as a GBL regulatory cassette. Through quorum sensing antibiotic production generally starts in early growth phase; a basal level of intracellular GBL is maintained throughout the lifetime of the cell, with derepression of ArpA-subfamily repressors only occurring once GBL concentrations are above a critical level[81]. Above this critical level derepression of ArpA-subfamily repressors

is self-sustained due to the expression of the GBL biosynthetic genes, which further produce more GBLs. These effectors are often referred to as autoregulators, and by extension, the operators which ArpA-subfamily repressors bind to are defined as Autoregulator-Responsive Elements (AREs).

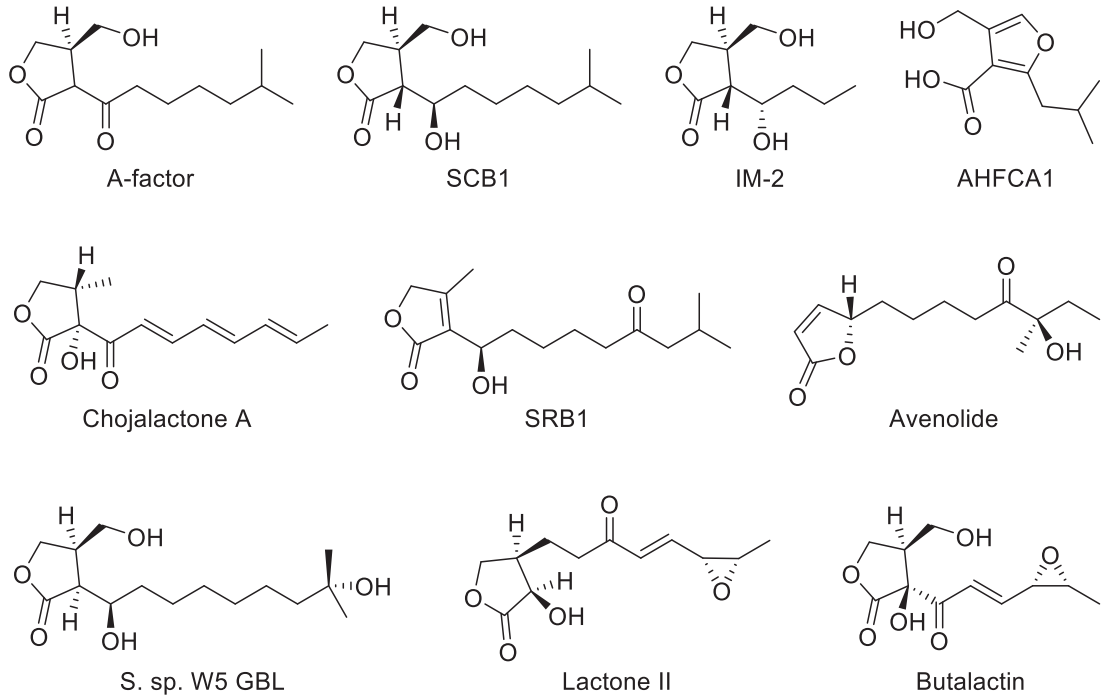
Despite being the namesake of the family, *S. griseus* ArpA is a less typical example of an ArpA-subfamily repressor in that the *arpA* gene is not in proximity to either the target of ArpA or the GBL biosynthetic genes. ArpA is not a direct pleiotropic regulator, but through its single target, *adpA*, which encodes an AraC/XylS-family pleiotropic transcriptional activator, it controls antibiotic biosynthesis and sporulation in *S. griseus*[82]. The biosynthetic gene cluster of streptomycin was characterized long before the sequencing of the *S. griseus* genome in 2008[83]; the streptomycin biosynthetic gene cluster was partially sequenced using self-cloning by Ohnuki *et al* in 1985[84], and all genes identified in 1987[85, 86]. The GBL signalling hormone A-factor (2-isocapryloyl-3*R*-hydroxymethyl-γ-butyrolactone) was shown to induce the biosynthesis of streptomycin in 1973[87], but it was not until 1989 that radioactively labelled A-factor was used to identify the target as the ArpA-subfamily repressor[88]. In 1991 the activity of an A-factor-dependent promoter upstream of the streptomycin gene cluster transcriptional activator *strR* was identified[89], which was later shown to be bound by AdpA 2 years later in 1993[90]. It was only comparatively recently in 1999 that ArpA was shown to repress streptomycin biosynthesis through repression of the *adpA*[91].

Since the characterisation of A-factor and its receptor, the number of known GBL signalling molecules has increased greatly (Figure 1-5). To date, there are three canonical subclasses of GBL signalling molecules; the butanolides, the 2-alkyl-4-hydroxymethylfuran-3-carboxylic acids (AHFCAs), and the butenolides. The butanolide subclass is further split into three types; the 6-keto-type, 6*R*-hydroxy-type, and 6*S*-hydroxy-type butanolides, the latter two being derived from 6-keto-type GBLs. All three butanolide types typically have a 2*R*, 3*R* stereochemistry of the lactone ring (with exceptions, such as *S.* sp. W5 GBL). ArpA-subfamily repressors receptive to one of these γ-butanolide types are often not sensing for other types, indicating that the 6-keto group and the stereochemistry of its reduced form are important in GBL recognition. Only two butenolide types have so far been characterised. *S. rochei* butenolides SRB1 and SRB2 which regulate lankacidin and lankamycin biosynthesis in *S. rochei*[92] are one type of butenolide with a similar biosynthetic origin to GBL and AHFCAs. Avenolide, which regulates avermectin biosynthesis in *S. avermitilis*[93], has a different ring stereochemistry with an unusual arrangement of moieties around the ring, and has a distinct biosynthetic pathway different to that of other GBLs.

Despite at first glance being structurally dissimilar to GBL signalling molecules, the biosynthesis of AHFCAs is closely interrelated to that of the butanolides (Chapter 1.2.4). It is not uncommon to find that *Streptomyces* species containing a GBL regulatory cassette also contain an AHFCA regulatory cassette. The most studied of these AHFCA regulatory cassettes is found control the biosynthesis of the peptidoglycan cell wall biosynthesis inhibitor methylenomycin A (MmA)[94], a on the *S. coelicolor* A3(2) plasmid SCP1[95] and the *S. violaceoruber* SANK-95570 plasmid pSV1[96].

In many *Streptomyces* spp. regulation by GBL/AHFCA/ regulatory cassettes are complicated by the presence of additional ArpA-subfamily repressors typically encoded in proximity to the GBL/AHFCA receptive regulator. The DBD of these additional repressors bear strong sequence identity with the other ArpA-subfamily repressor and often bind to many of the same ARE sequences. However, poor homology of the LBD results in non-recognition of the GBL-type ligand. These are referred to as pseudo-GBL receptors, or pseudo-AHFCA in the case of AHFCA-regulated systems. Although the literature suggests that the pseudo-GBL receptors are responsive to both endogenous and exogenous antibiotics[97], the role of many of these pseudo-AHFCA receptors is not yet understood.

In *S. coelicolor* A3(2) there are at least three ArpA-type regulatory systems functioning as pairs of repressors; CprA/CprB, ScbR/ScbR2 and MmfR/MmyR. CprB, in tandem with its 91 % sequence identity homologue CprA, are as previously described pleiotropic regulators of secondary metabolite biosynthesis and morphological differentiation[68]. However, unlike the other two systems these repressors are not thought to have a dichotomy of GBL and pseudo-GBL receptive roles, but instead may respond to the same currently unidentified GBL-type ligand (based on high sequence identity). The genes for ScbR and ScbR2 are localised in the *cpk* gene cluster where they directly regulate the production of CPK[98, 99]. ScbR is the receptor for the 6*R*-hydroxy-type *S. coelicolor* A3(2) butanolides (SCBs), and ScbR2 is the pseudo-GBL receptor. ScbR2 has been shown to be responsive to endogenous ACT and RED, coordinating biosynthesis of CPK with these other biosynthetic systems, as well as being responsive to the exogenous antibiotic jadomycin B (JdB) produced by the GBL-regulated *jad* gene cluster in *Streptomyces venezuelae*[100]. MmfR and MmyR are the regulators of the *mmy* AHFCA-regulated gene cluster, with MmfR functioning as the AHFCA receptor and MmyR the pseudo-AHFCA receptor. These last regulators, and the gene clusters they regulate, are one of the primary focuses of research done by the Corre group. Through examples of pleiotropic cross-talk between gene clusters described above, it is believed that these three *S. coelicolor* A3(2) systems may coordinate with one another to regulate the biosynthetic pathways of multiple antibiotic gene clusters through quorum sensing.



Organism	Effector biosynthetic enzymes	Effector	Receptor	Regulated Secondary metabolite	References
Butanolide-type					
S. griseus	AfsA, BprA	A-factor	ArpA	Streptomycin	(90, 101)
S. coelicolor A3(2)	ScbA, ScbB, ScbC	SCB1-3	ScbR	Coelimycin P1, Actinorhodin, Undecylprodigiosin	(66, 102)
S. venezuelae	JadW1, JadW2, JadW3	SVB1 (SCB3)	JadR3	Jadomycin B	(103)
S. virginiae	BarX, ?, BarS1	VB A-E	BarA	Virginiamycin	(104, 105)
S. lavendulae	FarX, ?, ?	IM-2	FarA	Showdomycin	(106, 107)
S. sp. CJ-5	unknown	Chojalactone A-C	unknown	unknown	(108)
S. sp. W5	unknown	Unassigned			
S. sp*., Gö* 40/10	unknown	Lactone (I & II) and Butalactin			(109)
AHFCA-type					
S. coelicolor A3(2)	MmfL, MmfH, MmfP	AHFCA(1-5)	MmfR	Methylenomycin	(95)
S. venezuelae	SgnL, SgnH, SgnP	AHFCA(5-8)	SgnR	Gaburedins	(110, 111)
S. avermitilis	AvaA, AvaB, AvaC	AHFCA(3, 5-7)	AvaL1	Unassigned azoxy compound	(112), this work
S. sclerotialus	*SclM2*, *SclM3*, *SclM5*	AHFCA (?)	SclM1	Scleric acid	(113), this work
Butenolide-type					
S. avermitilis	Aco, Cyp17	Avenolide	AvaR1, AvaR2	Avermectin	(77, 78, 93)
S. rochei	SrrX, CypX	SRB1 and SRB2	SrrA	Lankamycin and Lankacidin	(92)
S. fradiae	Orf18*, Orf16	Unknown	TylP	Tylosin	(114)
S. ghanaensis	Ssfg_07849, Ssfg_07847	Unknown	$AvaR_{gh}$	Moenomycin	(115)
S. griseoauranticus	Sgm_6045, Sgm_6043	Unknown	Sgm_6044	unknown	(93)
Uncharacterised / Other					
S. tsukubaensis	bulS1, bulS2	unknown	BulR1	Tacrolimus	(116)
S. acidiscabies	SabA	unknown	SabR	WS5995B	(117)
S. fradiae	AfsA-g	A-factor-like	ArpA-g1, ArpA-g2	Neomycin	(118)
S. SBI034	StcA, ?, ?		StcA	Bafilomycin	(119)
S. hygroscopicus	unknown	1,4-butyrolactone	ShbR3	Validamycin	(120)

Figure 1-5 – Examples of known effectors of *Streptomyces* ArpA-subfamily repressors. Adapted from the PhD thesis of Dr Sidda[111].

1.2.3 Cooccurring GBL/AHFCA-regulatory cassettes control individual antibiotic pathways

Streptomyces containing AHFCA-regulated gene clusters also frequently encode one or more GBL-regulated gene clusters (Figure 1-6). As mentioned, *S. coelicolor* A3(2) contains three systems of ArpA regulators, but the biosynthetic genes for GBL and AHFCA biosynthesis are encoded within the *cpk* and *mmy* gene clusters respectively. Other AHFCA regulatory systems which are currently being investigated by the Corre group include the gaburedin (*gbn*) biosynthetic gene cluster in *S. venezuelae* ATCC 10712, an azoxy compound biosynthetic gene cluster in *S. avermitilis* MA-4680, and the scleric acid (*scl*) biosynthetic gene cluster in *Streptomyces sclerotialus* NRRL ISP-5269[110, 112, 113]. These gene clusters were identified using bioinformatic BLAST analysis using regulator and AHFCA biosynthetic genes as queries, and boundaries of the gene clusters predicted using antiSMASH. AHFCA regulatory cassettes have also been identified in *Streptomyces hygroscopicus* subsp. *jinggangensis* 5008, *Streptomyces* sp. HGB0020, *Streptomyces roseochromogenes* subsp. *Oscitans* DC 12.976, *Streptomyces iakyrus* NRRL ISP5482, *Streptomyces* sp. NRRL F-5135, *Streptomyces* sp. S-31, and *Streptomyces roseoverticillatus* NRRL B-3500, but have yet to be investigated. In addition, an AHFCA regulatory cassette has been identified in the actinomycetes bacterium *Kitasatospora cheerisanensis* KCTC 2395, suggesting that AHFCAs as a signalling molecule may not be unique to the *Streptomyces* genus. This in turn could be indicative of a shared evolutionary trait, or example of HGT between bacteria from the different genera.

Deletion of ArpA-subfamily repressors is a strategy that has already been employed to great effect in unlocking the cryptic gene clusters of novel secondary metabolites; the deletion of the pseudo-GBL receptor *scbR2* in *S. coelicolor* A3(2) led to the characterisation of coelimycin[121] while deletion of the pseudo-AHFCA receptors Δ*gbnR*, Δ*aval2* and Δ*sclM4* in *S. venezuelae*, *S. avermitilis* (heterologous host *S. lividans* TK24 pESAC13A-2H1) and *S. sclerotialus* respectively have led to the discovery of the metabolites produced by those AHFCA-regulated gene clusters[111-113]. However, deletion of the GBL/AHFCA receptors often does not unlock the production of these natural products; Δ*mmyR* results in constitutive overexpression of MmA and AHFCAs, while Δ*mmfR* results in less MmA being produced than the wild type[67], and similarly Δs*cbR* does not result in production of either coelimycin or SCBs[98].

S. venezuelae, originally isolated from Caracas, Venezuela in 1947 and named in 1948, is amongst the early antibiotic producers investigated in the late 1940s. It is noted to produce the antibiotic chloramphenicol, sold under the trade name Chloromycetin[122, 123]. The *S.*

venezuelae ATCC 10712 genome, sequenced in 2011[124], has a size of 8.23-Mb with 72.4 % GC content, encoding 7,500 genes on a single linear chromosome. The products of the AHFCA-regulated gene cluster, the gaburedins, are relatively simple molecules comprised of γ-aminobutyrate (GABA; produced by decarboxylation of the amino acid glutamate) joined *via* a urea bond to amino acids with hydrophobic side-chains which include phenylalanine, leucine, isoleucine and valine. The activity of gaburedins are not known, but comparison with that of other urea-containing compounds such as syringolin, MG132 and the pacidamycins suggest that gaburedins may function as protease inhibitors[125, 126]. The *jad* gene cluster in *S. venezuelae* is regulated by the GBL receptor JadR3 and the pseudo-GBL receptor JadR2[127, 128]. Δ*jadR3* results in a decrease of JdB production[103], while Δ*jadR2* ethanol induction-dependent increases JdB production[97]. Only one GBL signalling molecule has been observed in *S. venezuelae* thus far; a 6*R*-hydroxy-type GBL named *S. venezuela*e butanolide 1 (SVB1), which has an identical structure to SCB3 produced by *S. coelicolor* A3(2), indicating a shared biosynthetic pathway[103] and the possibility of these species communicating with one another *via* quorum sensing. Meanwhile, there is evidence that the effectors of JadR2 are the endogenous antibiotics JdB and chloramphenicol, forming part of a feed-forward link that coordinates the two biosynthetic systems[97].

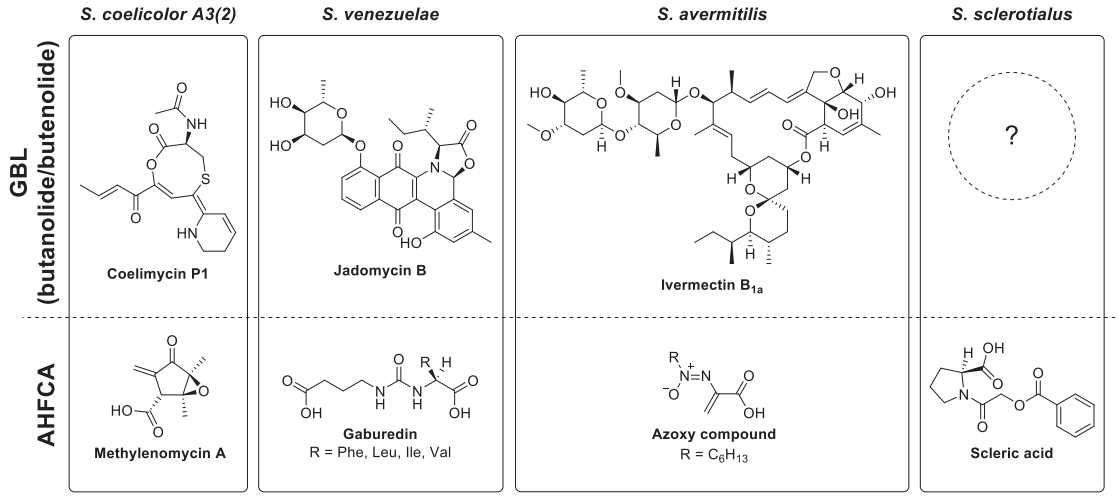


Figure 1-6 – Products of secondary metabolite biosynthetic gene clusters regulated by either GBL or AHFCA autoregulatory cassettes.

The *ave* gene cluster in *S. avermitilis* is regulated by the ArpA regulators (AvaR1, AvaR2 and AvaR3), and is induced in the presence of avenolide (discussed in Chapter 1.3.5)[45, 93]. The AHFCA-regulated gene cluster has high gene homology with the valanimycin gene cluster of *Streptomyces viridifaciens*, which produces an azoxy compound *via* involvement of an unusual seryl-tRNA synthetase[112]. Δ*avaL2* was previously attempted in *S. avermitilis* but proved to have poor genetic tractability (but not complete intractable, as other mutations in *S. avermitilis* have been performed). Instead, a pESAC vector containing the biosynthetic gene

cluster was studied in the heterologous conjugant *Streptomyces lividans* TK24 pESAC13A-2H1[112]. Δ*avaL2* in this host resulted in the production of a yet-to-be assigned azoxy compound structurally similar to that of valanimycin. Some azoxy-containing compounds are known for their highly genotoxic behaviour, being frequently carcinogenic[129, 130], but they also have applications as antimicrobial, anticancer and antifungal drugs[131]. Crude extracts of *S. lividans* TK24 pESAC13A-2H1 Δ*avaL2* had observable activity against *Bacillus subtilis* and *Staphylococcus aureus*, as well as against wild type *S. lividans* TK24, indicating that the gene cluster also encodes the azoxy compound resistance gene.

S. sclerotialus is a poorly characterised thermophilic bacterium named for its ability to form sclerotia, a hard mass of fungal or fungal-like mycelium containing food reserves which can ensure survival in a dormant state for extended periods of time. It was originally isolated in 1955 from thermal springs in Poona, India as a member of the *Actinomycetales* species C*hainia antibiotica* Thirumalachar[132], and then later reclassified into the *Streptomyces* genus[133]. Only partial shotgun sequences of the genome currently exist for two strains, *S. sclerotialus* NRRL B-2317 and *S. sclerotialus* NRRL ISP-5269, submitted by the University of Illinois as part of a large-scale metagenomics study investigating discovery of phosphonic acid natural products[134]. *S. sclerotialus* NRRL ISP-5269 encodes an AHFCA regulatory cassette in a gene cluster found on contig 43 that has been shown to produce the antibiotic scleric acid[113]. Scleric acid exhibits moderate activity against *Mycobacterium tuberculosis* and inhibits nicotinamide N-methyltransferase (NNMT), a metabolic enzyme associated with cancer. No GBL gene cluster has currently been characterised but blasting *S. sclerotialus* NRRL ISP-5269 with ArpA shows two additional ArpA homologues on contig 28 (WP_030622529.1, 45 % sequence identity and WP_051872298.1, 29 % sequence identity), which in antiSMASH are part of a gene cluster that also contains GBL biosynthetic genes. Partial alignments of an ArpA-like DBD are also found on contig 24 (WP_037773057.1, 36 % sequence identity). It is not known whether either of these GBL/AHFCA systems are encoded within a plasmid or integrated into the *S. sclerotialus* chromosome.

The relative prevalence of GBL-regulated gene clusters occurring across *Streptomyces* species, compared to AHFCA-regulated gene clusters suggests that the appearance of GBL-regulated systems may be an earlier evolutionary event. One probable scenario is that, given evidence of the cassette on the *S. coelicolor* A3(2) SCP1 plasmid, the regulatory components themselves likely evolved by paralogous gene duplication of GBL ArpA receptors and biosynthetic proteins, and subsequently have transferred horizontally between species on a mobile DNA element, the cassette having differentiated to perform its AHFCA responsive function prior to HGT. The absence of identical gene clusters being retained between species

after HGT is likely explained if the cassette originally existed independent of any biosynthetic gene cluster, much like other GBL-regulated systems such as ArpA. Therefore, in its original form this likely played a pleiotropic role that, when integrated into the host's DNA incorporated itself well into the pre-existing GBL-regulated pathways. *S. coelicolor* A3(2) may be an example of this process in an intermediary phase, with the *mmy* gene cluster having not yet integrated into the chromosome. GBL-regulated gene clusters also share heterogeneity of gene organisation in their gene clusters, suggesting they themselves might have a similar origin. It is not known whether pleiotropic or pathway specific activity came before the other, or if they co-evolved together. Alternatively, an earlier GBL gene-duplication event may have occurred, and both GBL and AHFCA systems may have coevolved across species.

1.2.4 Biosynthesis of GBL/AHFCA signalling molecules

While the biosynthetic route of more complex GBL signalling molecules, such as chojalactones or butalactin have not been investigated, the biosynthesis of simpler 6-keto-type, 6*R*-hydroxy-type, 6*S*-hydroxy-type and AHFCA signalling molecules with *n*-alkyl, isoalkyl or *sec*-alkyl moieties has been well characterised. In particular, work done in *S. coelicolor* A3(2) and *S. venezuelae* by the Corre group have shown that all these GBL and AHFCA signalling molecules have a shared biosynthetic pathway(Figure 1-7)[111, 135].

The biosynthesis of the 6-keto-type butanolides is best studied in the biosynthesis of A-factor in *S. griseus.* The biosynthesis of A-factor is catalysed by the GBL-biosynthesis enzyme AfsA[101, 136] and the butenolide phosphate reductase BprA[137], which are encoded in the same operon. In other ArpA systems the *afsA-bprA* operon is often under the direct control of the ArpA-homologue, however studies have shown that *afsA* transcription is not regulated by ArpA in *S. griseus*[138]. AfsA catalyses the esterification of dihydroxyacetone phosphate (DHAP) and a β-ketothioester linked to either acyl carrier proteins (ACP) or coenzyme A (CoA). DHAP is generated either through the fructose-bisphosphate aldolase-catalysed conversion of fructose 1,6-bisphosphate into glyceraldehyde 3-phosphate (G3P) and DHAP in the glycolysis pathway or converted directly from glycerol through glycerol kinase and glycerol 3-phosphate dehydrogenase. β-ketothioesters are generated through the fatty acid metabolism pathway using acyl-CoA starter units, which determine the alkyl moiety of the signalling molecule (Figure 1-7). As *Streptomyces* produces both linear and branched alkyl fatty acids, the signalling molecules are also either linear or branched. After condensation, the intermediate undergoes a spontaneous intramolecular aldol reaction to form a phosphorylated 6-keto-type butenolide intermediate.

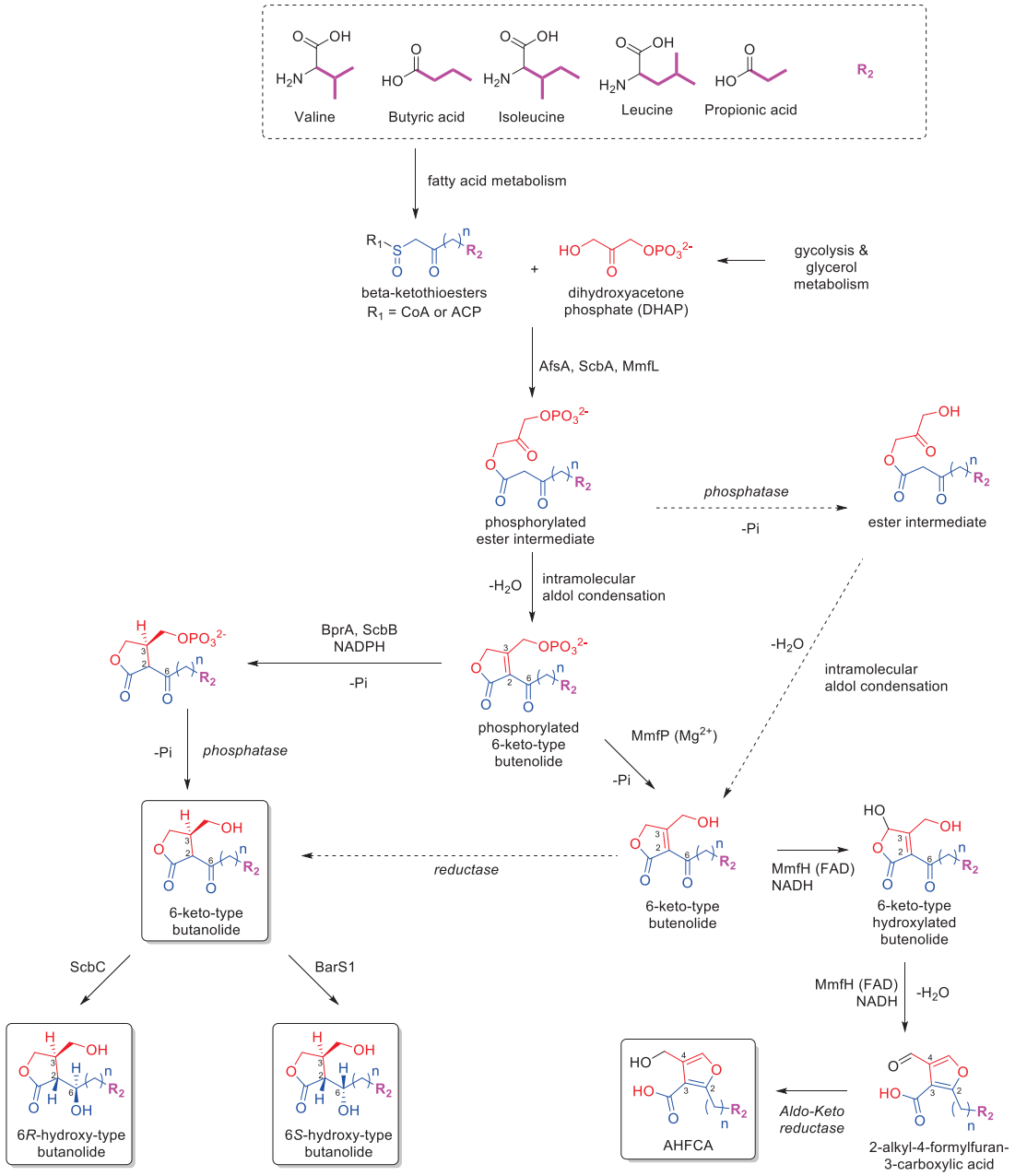


Figure 1-7 – Proposed shared biosynthetic pathway for 6-keto-type, 6*R*-hydroxy-type and 6*S*-hydroxy-type 2*R*,3*R* GBL and AHFCA signalling molecules. Adapted from biosynthesis pathways described in the theses of Dr Sidda and Dr Malet[111, 139]. Formation of a shared 6-keto-type butenolide is catalysed by AfsA homologues (*e.g.,* ScbA, MmfL), which is required for the biosynthesis of both GBL and AHFCA signalling molecules. Primary pathways are indicated by solid arrows, and final autoregulatory molecules are boxed. Enzymes in italics are not currently identified.

BprA was initially thought to catalyse both the dephosphorylation of the 3-hydroxylmethyl moiety and reduction of the butenolide, forming the 6-keto-type butanolide with 2*R*, 3*R* alkyl stereochemistry. However, Kato *et al.* later showed that BprA has only reductase activity, and the dephosphorylation step to form A-factor was catalysed by a different phosphatase[137]. A similar pathway was previously proposed for the biosynthesis of *Streptomyces virginiae* butanolide (VB) A[140]. An alternative pathway was also demonstrated where the ester intermediate undergoes dephosphorylation by another phosphatase prior to aldol

condensation, and then was reduced by a different reductase to form A-factor[137]. These phosphatase and reductase enzymes were commonly present bacterial enzymes not specific to the A-factor pathway. Because the *afsA-bprA* is essential for streptomycin production, the BprA-dependent pathway is thought to be the major route for A-factor biosynthesis. The 6*R*-hydroxy-type and 6*S*-hydroxy type butanolides are derived from the 6-keto-type butanolide by action of an additional stereospecific reductase in *S. coelicolor* this is *SCO6264* encoded convergently adjacent to *scbR* (*SCO6265*), and *barS1* in S. *virginiae* is similarly nearby to other GBL-regulatory genes[141]. In *S. coelicolor* A3(2) AfsA and BprA homologues are encoded divergently from the *scbR* gene in the same operon; *scbA* (*SCO6266*) and *scbB* (*SCO6267*) respectively. In the literature *scbB* is sometimes erroneously labelled as *scbC*, which is here attributed to *SCO6264*.

As mention prior, the biosynthesis of butenolides may or may not be directed through different biosynthetic machineries. The butenolides SRB1 and SRB2 are produced through a presumably similar mechanism as the butanolides, as an AfsA homologue, SrrX, is encoded in the GBL regulatory cassette of *S. rochei*[142]. In the case of avenolide however, biosynthesis is catalysed by a different system comprised of an acyl-CoA oxidase, Aco, and a cytochrome P450 hydroxylase, Cyp17[93]. The mechanism of biosynthesis for this type of butenolides is not yet understood. Homologues of the Aco/Cyp17 system are also found in *S. fradiae* (regulator TylP[114, 143]), *S. ghanaensis*, and *S. griseoauranticus*. In these systems, *cyp17* is encoded in the same operon as the ArpA-subfamily repressor, while *aco* is divergent.

The biosynthetic genes of AHFCA signalling molecules, much like the GBL systems, encode a triad of catalytic enzymes. In *S. coelicolor* A3(2) these enzymes are MmfL, an AfsA homologue, MmfH, a flavin-dependent oxidoreductase, and MmfP, a haloacid dehalogenase (HAD)-like phosphatase, which are divergently adjacent from the AHFCA receptor mmfR encoded in a single operon (*mmfLHP*). In the *Streptomyces* species being investigated the *mmfR* and *mmyR*-type regulator are always regulated separately, and *mmfL* homologues are always regulated at the same site divergent from MmfR, but the *mmfH*-type and *mmfP*-type ORFs are positioned in different operons. The biological significance of this arrangement, if any, is not understood. Like ScbA, MmfL is necessary for AHFCA biosynthesis as evidenced by transformation of *mmfL* into *E. coli* resulting in production of AHFCAs[95], as well as the formation of a ^{13}C-labelled phosphorylated 6-keto-type GBL after feeding with glycerol[144]. *In vitro* work performed by Dr Malet suggests that MmfP acts first upon the phosphorylated 6-keto butenolide intermediate by catalysing the Mg^{2+} dependent dephosphorylation of the 3-methylphosphate to 3-methylhydroxy without hydrogenation of the lactone ring, forming a 6-keto-type butenolide intermediate like that proposed in the alternate A-factor biosynthetic

pathway[139]. MmfH was also shown to be flavin adenine dinucleotide (FAD) dependent, requiring nicotinamide adenine dinucleotide (NADH) presumably to reduce FAD to $FADH_2$, however the mechanism is not yet understood. What is proposed is a two-step reaction that first involves hydroxylation of the lactone ring, followed by subsequent rearrangement and dehydration to form the furan ring. The resulting compound was not the final AHFCA, but instead 2-alkyl-4-formylfuran-3-carboxylic acid, containing methyl aldehyde instead of a hydroxymethyl moiety. Reduction of the aldehyde is proposed to be carried out by a non-specific reductase to yield the final AHFCA product.

It has been demonstrated by Dr Sidda that both GBLs and AHFCA produced in *S. venezuelae* share the same fatty acid metabolism precursors[111]. By extension, this is likely the case for other *Streptomyces* spp. as well. *S. coelicolor* A3(2) naturally produces the GBLs SCB1-3, as well as AHFCA1-5 (Table 1-1). While the same fatty acid metabolic precursors are utilised by both biosynthetic systems, there is no overlap in the utilisation of the resulting β-ketothioesters that are incorporated into each GBL/AHFCA molecule. This shows that the biosynthetic enzymes, presumably ScbA/MmfL, are selecting for specific β-ketothioester precursors. The *S. coelicolor* A3(2) mutant strain *S. coelicolor* M1152 (Δ*act* Δ*red* Δ*cpk* Δ*cda rpoB[C1298T]*), which retained the SCB regulatory cassette (with the exception of *scbR2*) produced the additional molecules SCB4-8, however this placed metabolic strain upon the organism, reducing the yield of the coelimycin[145]. In addition, some of these additional SCBs showed significantly weaker activity in releasing ScbR from DNA compared with SCB1-3, suggesting that perhaps the GBL/AHFCA biosynthetic machinery in each organism is optimised for maximum ligand efficacy against the target regulator, or vice versa.

Table 1-1 – Shared biosynthetic precursors of SCBs produced by ScbA and AHFCAs produced by MmfL in *S. coelicolor* A3(2). SCBs in bold are produced in the wild type *S. coelicolor* A3(2) strain. Other SCBs are overproduced in the *S. coelicolor* M512 superhost strain (Δ*act* Δ*red* Δ*cpk* Δ*cda rpoB[C1298T]*). AHFCAs in bold were observed in the *S. coelicolor* M512 pSET152:*mmfLHP*. Other AHFCA are observed in genetic and precursor studies in *S. venezuelae* and transformed *E. coli*[111]. Number of carbons added in chain extension during fatty acid biosynthesis is shown in parenthesis. * SCB9 and SCB10 are theoretical isomers based on mass fragments of produced SCBs with m/z = 281 and 239 respectively.

Alkyl Moiety	Alkyl mW (Da)	Precursor	SCB	AHFCA
propyl	43	?	**-**	**AHFCA2**
Isobutyl (2-methylpropyl)	57	?	**-**	**AHFCA1**
butyl	57	?	**-**	**AHFCA4**
Isopentyl (3-methylbutyl)	71	Valine (1)	SCB10*	**AHFCA3**
Pentyl	71	Butyric acid (1)	SCB8	**AHFCA5**
3-methylpentyl	85	Isoleucine (1)	SCB4	-
Isohexyl (4-methylpentyl)	85	Leucine (1)	SCB5	AHFCA6
hexyl	85	Propionic acid (3)	SCB6	AHFCA7
Isoheptyl (5-methylhexyl)	99	Valine (3)	**SCB1**	AHFCA8
Heptyl	99	Butyric acid (3)	**SCB2**	AHFCA8E
5-methylheptyl	113	Isoleucine (3)	**SCB3**	AHFCA9
Isooctyl (6-methylheptyl)	113	Leucine (3)	SCB9*	AHFCA10
octyl	113	Propionic acid (5)	SCB7	-

Also, of note is that the transformation of the *afsA-type* genes alone into *E. coli* is sufficient for the production of either 6-keto-type GBLs or AHFCAs. However, this is further complicated in that *mmfL*-type transformations do not result in the production of 6-keto-type GBLs as seen in *afsA* transformations[137], and vice versa. For AfsA this was rationalised by the presence of a BprA phosphate reductase homologue in *E. coli*; quinone reductase 2 (WP_016159986.1), which performs a similar role. An MmfP homologue is also found in *E. coli* (WP_096136332.1, HAD-family hydrolase, 29 % sequence identity), as well as MmfH homologue, acyl-CoA dehydrogenase (WP_045396622.1, 26 % sequence identity). Therefore, it is possible that, like the observed alternative pathway for A-factor biosynthesis, there is also sufficient homology of specific or non-specific bacterial reductase/phosphatase enzymes to perform the roles of MmfP and MmfH.

No 6*R/S*-hydroxy-type GBLs were observed for *E. coli afsA* transformants even though there are numerous ScbC homologues encoded in the *E. coli* genome (highest identity is SDR-family oxidoreductase, WP_003023319.1, with 40 % sequence identity). A recent paper by Takano *et al.* showed that transformation of *scbABC* into *E. coli* results in the production of SCB2[146]. The *scbAB* transformant (*scbAC* in the paper) resulted in the production of the 6-keto-type GBL, consistent with the homologous activity of AfsA and BprA (although the necessity of *scbB* is questionable since A-factor is also produced when only *afsA* is transformed). The *scbAC* transformant (*scbAB* in the paper) resulted in the production of a novel dephosphorylated 6*R*-hydroxy-type butenolide. This showed that ScbB is required for the formation of the 6*R*-hydroxy-type butanolide, and that ScbC can function *via* an ScbB independent pathway. There is currently no evidence of *scbA* transformed in isolation into *E. coli* producing linear alkyl-chained SCB analogues.

Deletion of the pseudo-AHFCA receptors in the *S. coelicolor* system reveals that different *Streptomyces* species also produce different AHFCA signalling molecules (Figure 1-8). Δ*avaL2* in *S. lividans* TK24 pESAC13A-2H1 resulted in production of AHFCAs 3 and 5-7, while *S. venezuelae* Δ*gbnR* produces AHFCAs 5-8, with AHFCA9 and AHFCA10 observed when fed with d10-labelled leucine and isoleucine fatty acid precursors[111, 112]. The transformation of *S. coelicolor* A3(2) *mmfL* and *S. venezuelae sgnL* into *E. coli* showed that differences in produced AHFCA are influenced by the AfsA homologue enzyme, as the *E. coli sgnL* transformant produced a larger novel linear AHFCA8 isomer, AHFCA8E[111]. It is thought that the selectivity for β-ketothioesters recognised by MmfL-type enzymes has coevolved for the interaction with that specific AHFCA receptor. Thus, MmfR-type receptors from different species are expected to exhibit different AHFCA recognition profiles.

However, the selectivity of the AHFCA receptors has only been studied in MmfR, while SgnR is the only other repressor confirmed to be responsive to AHFCAs.

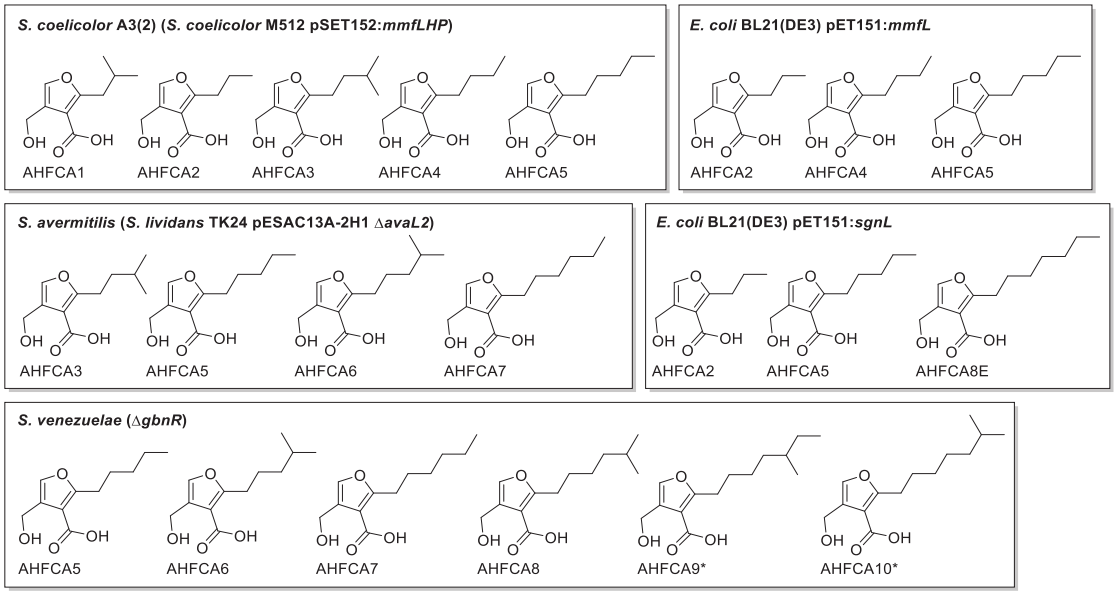


Figure 1-8 – Structures of AHFCA molecules produced *in vivo* by *Streptomyces* spp. and *E. coli* transformed with AHFCA biosynthetic genes. Asterisks Indicate novel AHFCAs produced in feeding studies with d10-labelled leucine and isoleucine.

1.3 AHFCA-regulated gene clusters in *Streptomyces*

1.3.1 The *S. coelicolor* A3(2) methylenomycin (*mmy*) gene cluster

The mmy gene cluster contained on the SCP1 plasmid encodes 21 genes (Figure 1-9A). Currently the *mmy* gene cluster is the only known AHFCA-regulated gene cluster which is not found in the linear chromosome. In addition to the two ArpA-subfamily repressors, the gene cluster encodes two other transcriptional regulators; *mmyB* and *mmyJ*. MmyB is the transcriptional activator for MmA biosynthesis, which is proposed to recruit RNAP to palindromic B-box sites upstream of the *mmyYF*, *mmyBQEDXCAPK*, and *mmyTOG* biosynthetic operons[67]. MmyJ is an ArsR-family regulator that controls the expression of neighbouring *mmr*, encoding a putative transmembrane MmA efflux pump. MmA binds to and releases MmyJ from the *mmr* promoter, demonstrating the role of MmyJ as the MmA resistance regulator[147]. *mmyB* and *mmfL* both contain a TTA codon, therefore successful translation is dependent on accumulation of BldA, consistent with the production of MmA in late exponential-stationary phase growth[67].

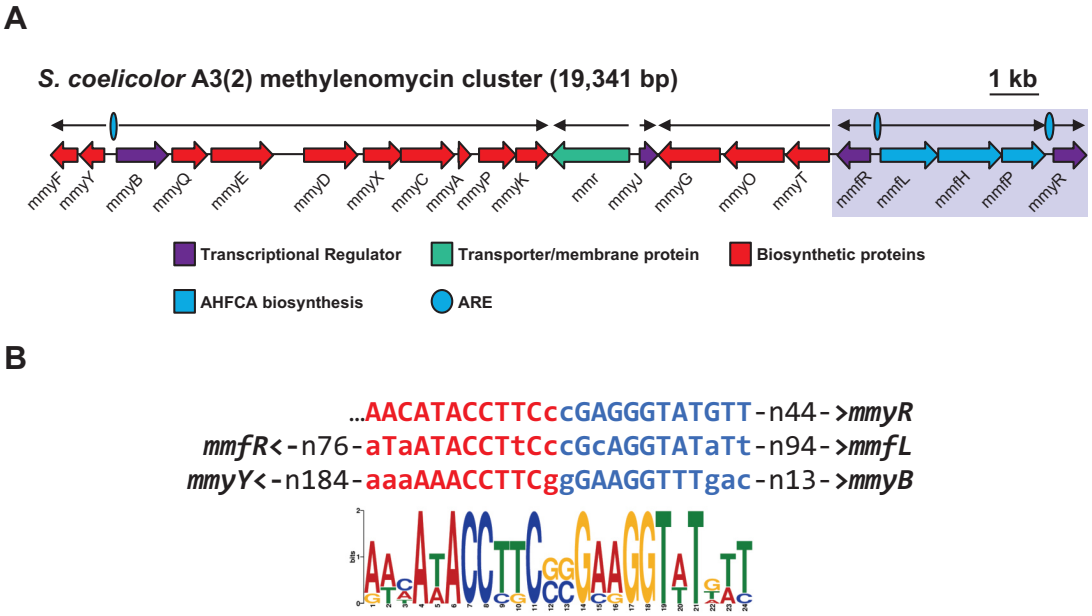


Figure 1-9 – Genetic organisation of the AHFCA-regulated methylenomycin (*mmy*) gene cluster encoded in the SCP1 plasmid of *S. coelicolor* A3(2). (**A**) Annotated map of genes, operons and confirmed AREs bound by the AHFCA receptor MmfR. The AHFCA regulatory cassette is enclosed in a blue box. Putative functions of *mmy* genes are listed in Table 8-1, Appendix A. (**B**) Sequences of confirmed pseudo-palindromic AREs bound by MmfR. Half-dyads around the dyad centre are coloured **red** and **blue** respectively; nucleotides in upper case are palindromic around the dyad centre. n indicates the distance in nucleotides to the indicated ORF start codon. The MEME palindromic motif is shown underneath.

MmfR binds to three ARE sites within the gene cluster (Figure 1-9B), two of which are found within the AHFCA regulatory cassette[139, 148]. The first site is upstream of its own gene *mmfR* and the divergent *mmfLHP* operon encoding the AHFCA biosynthetic enzymes (*mmfR/mmfLHP*$_{ARE}$), overlapping the bidirectional promoters for these operons. The other site within the regulatory cassette is found upstream of *mmyR* (*mmyR*$_{ARE}$); thus, all genes within the regulatory cassette are repressed by *MmfR*. The final site is found upstream of *mmyB* downstream of the proposed *mmyB* B-box (*mmyB*$_{ARE}$), thereby controlling expression of both the activator and the downstream biosynthetic genes. MmfR likely does not control the transcription of *mmyYF*; the divergent *mmyY* start codon is 184-bp downstream of the ARE, upstream of proposed B-boxes in this region. The AREs to which MmfR binds have a minimal length of 22-bp to 24-bp, with AT-rich ends at each end of the sequence, and CG rich towards the middle, except for a TT/AA rich region 3 bp to 4 bp from the dyad centre. The nucleotides 8-bp to 11-bp from each side of the dyad centre are commonly referred to as TATA box-like regions, in reference to the promoter Pribbenow boxes with which these operators commonly overlap with. Because of the palindromic nature of the ARE the strand orientation is irrelevant to the direction in which transcription is repressed, and thus binding of MmfR to these AREs has bidirectional repressive activity, such as in the case of the *mmfR*/*mmfLHP* region.

MmyR is proposed to also bind these sites, based on homology of the DBD with MmfR. A bioluminescent reporter system was developed by Dr Styles using these ARE sites located

upstream of *lux* reporter genes contained on plasmids under the control of MmyR and MmfR transformed into *S. coelicolor* M145 (SCP1$^-$, SCP2$^-$)[149]. Her data showed significantly weaker binding of MmyR to all AREs compared to MmfR; MmfR binding at *mmfR*/*mmfLHP*$_{ARE}$ caused a 90 % reduction in luminescence compared with 32 % by MmyR. Other findings indicated that MmyR bound the *mmyB*$_{ARE}$ the strongest, and did not bind to its own operator, indicating it is likely not autoregulatory. These results also indicated that MmyR binding to the *mmyB*$_{ARE}$ increased in the presence of AHFCAs, although this is not yet explained. Attempts to produce recombinant MmyR *in vitro* have so far failed to yield soluble protein that can be assayed *in vitro*.

The AHFCAs produced by the methylenomycin system are specifically referred to as the methylenomycin furans (MMFs) in the context of *S. coelicolor* A3(2), but otherwise AHFCA is the preferred assignment. Both *in vivo* and *in vitro* assays of AHFCA activity against MmfR as well as the cocrystallisation of MmfR with AHFCA indicate that MmfR is the AHFCA receptor[139, 148]. The effector for MmyR, assuming there is one, is currently unknown, although based on homology it is probable that like other pseudo-GBL receptors MmyR is an antibiotic receptor.

1.3.2 The *Streptomyces venezuelae* gaburedin (*gbn*) gene cluster

The expression of biosynthetic genes in the AHFCA-regulated gene cluster of *S. venezuelae,* unlike that of *S. coelicolor* A3(2), are not under the control of a transcriptional activator. The *gbn* gene cluster is the simplest known AHFCA-regulated system, encoding a total of eight functional genes, including two smaller hypothetical proteins (Figure 1-10A). The AHFCA receptor SgnR, in addition to binding sites upstream of *sgnR*/*sgnL-SVEN_4184-sgnP* (*sgnR/sgnL-4184-sgnP*$_{ARE}$) and *SVEN_4186-gbnR*/*sgnH* (*gbnR-4186/sgnH*$_{ARE}$) within the regulatory cassette, has been shown to control the expression of the *gbnABC* operon which encodes two biosynthetic enzymes; GbnA (pyridoxal phosphate (PLP)-dependent glutamate decarboxylase) and GbnB (ATP-dependent acyl-CoA synthetase), as well as GbnC, an EamA-like membrane transport protein proposed as the gaburedin exporter[148]. The sequences of this ARE (*gbnABC*$_{ARE}$) and the other two sites are very similar to those found in the *mmy* gene cluster; so much so that they were initially identified by bioinformatic prediction and confirmed to be bound by SgnR using electrophoretic mobility shift assays (EMSA; alternatively gel shift assays). Gel shift assays also showed SgnR being released from DNA by AHFCAs, as has also been observed by induction *in vivo*[111, 148]. As stated before, Δ*gbnR* (pseudo-AHFCA receptor) results in constitutive expression of gaburedins, establishing that GbnR has a regulatory role in this gene cluster[110].

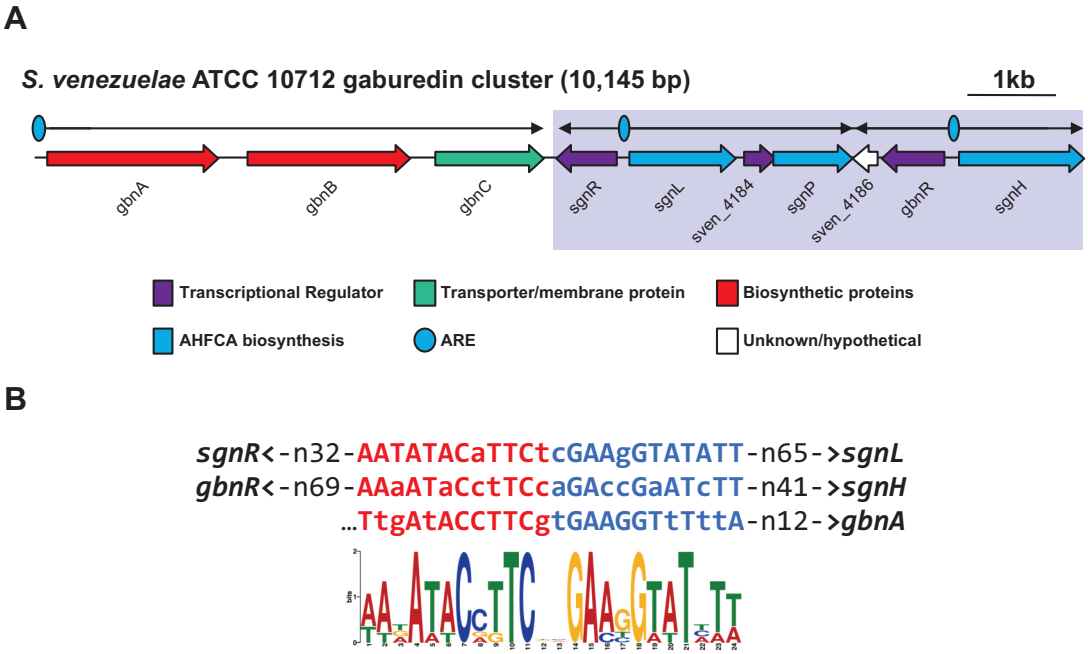


Figure 1-10 – Genetic organisation of AHFCA-regulated gaburedin (*gbn*) gene cluster in *S. venezuelae* ATCC 10712. (**A**) Annotated map of genes, operons and confirmed AREs bound by the AHFCA receptor SgnR. The AHFCA regulatory cassette is enclosed in a blue box. Putative functions of *gbn* genes are listed in Table 8-2, Appendix A. (**B**) confirmed AREs bound by SgnR. Half-dyads around the dyad centre are coloured **red** and **blue** respectively; nucleotides in upper case are palindromic around the dyad centre. n indicates the distance in nucleotides to the indicated ORF start codon. MEME palindromic motif is shown below.

SVEN_4184 encodes an uncharacterised partial LysR transcriptional regulator. LysR regulators are tetrameric dual repressor/activators usually comprised of ~300 amino acid residues which are, like TFRs, comprised of an N-terminal DBD and a C-terminal LBD[150]. SVEN_4184 by comparison has a length of 98 residues, encoding only the canonical DBD. SVEN_4184 has 28.0 % sequence identity with another truncated LysR regulator, *Pseudomonas putida F1* TodR, which showed no significant role in regulating neighbouring genes involved in the toluene degradation pathway[151]. It is possible that this truncated LysR is either non-functional or is part of a novel two-component signal transduction system. Alternatively, SVEN_4184 may simply permanently repress its target, being unresponsive to any effector. Although no analysis of the transcripts has been performed, TTA codons are found in the *SVEN_4184* (16-bp to 18-bp) and *mmfH*-like s*gnH* (829-bp to 831-bp) ORFs, but not in the *mmfL* homologue *sgnL*, indicating that BldA maturation is at least probably required by components of the AHFCA biosynthetic system, as well as *SVEN_4184*.

1.3.3 The *Streptomyces avermitilis* azoxy compound gene cluster

Of the AHFCA-regulated gene clusters characterised thus far, the 50,462-bp azoxy compound gene cluster found in *S. avermitilis* is the largest and most complex (Figure 1-11). It encodes 55 genes, including three additional transcriptional regulators; *saverm_2263* (TFR regulator), *saverm_2301* (SARP-family transcriptional regulator) and *saverm_2308* (PaaX-family

transcriptional regulator). Saverm_2301 is the proposed transcriptional activator for azoxy compound biosynthesis, suggested by Dr Poon to bind sites upstream of the four putative operons; *saverm_2282-saverm_2272*, *saverm_2283-saverm_2294*, *saverm_2300-saverm_2295* and *saverm_2301-saverm_2304*. Also encoded are four transporters; three MFS transporters (*saverm_2264*, *saverm_2275* and *saverm_2297*) and an amino acid permease (*saverm_2293*). Several membrane proteins with enzymatic activity are also predicted. Saverm_2263 is a putative type I TFR, proposed to regulate transcription of divergent *saverm_2264*, which is the putative azoxy-compound transporter. By extension, the azoxy compound is the proposed effector of Saverm_2263. However, there is no experimental evidence to yet confirm this functionality.

Only the pseudo-AHFCA receptor AvaL2 has been demonstrated thus far to have regulatory activity on this gene cluster; as for the *mmy* and *gbn* gene clusters, Δ*avaL2* results in constitutive overexpression of this azoxy compound, but the deletion of the AHFCA receptor AvaL1 results in no phenotypic change from wild type[112]. Similar to the *mmy* gene cluster, TTA codons are found in *mmfL*-like *avaA* (535-bp to 537-bp), *mmfH*-like *avaB* (316-bp to 318-bp) and *saverm_2301* (46-bp to 48-bp), indicating that production of the azoxy compound is likely dependent on BldA in a way comparable with the *mmy* gene cluster.

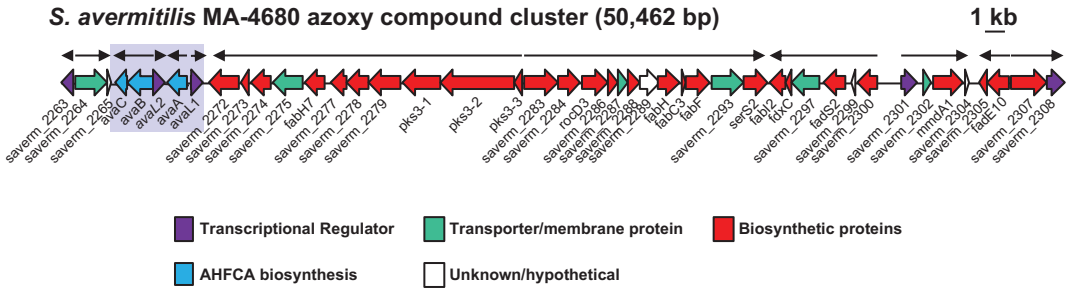


Figure 1-11 – Genetic organisation of the AHFCA-regulated cryptic azoxy compound gene cluster in *S. avermitilis* MA-4680. The AHFCA regulatory cassette is enclosed in a blue box. Putative functions of cluster genes are listed in Table 8-4, Appendix A. Putative operons are those proposed by Dr Poon[112].

1.3.4 The *Streptomyces sclerotialus* scleric acid (*scl*) gene cluster

The *S. sclerotialus* ISP-5269 *scl* gene cluster encodes 18 genes (Figure 1-12), which, like the *gbn* gene cluster, do not encode any other transcriptional regulators apart from a partial LysR repressor, SclL. Like SVEN_4184, SclL has an amino acid length of 109 residuals corresponding to the LysR DBD and has 20.7 % sequence identity with TodR. That these genes coincide in systems without any other transcriptional regulator may suggest that these two LysR regulators, if functional, perform similar roles in both systems, and as this gene is also present in the operon divergent from the gene encoding the MmfR-type repressor in both

systems, this may be evidence of a shared evolutionary ancestry. Unlike in the other AHFCA-regulated systems, no TTA codons are found in any of the genes within this gene cluster, possibly indicating that the gene cluster is either not directly dependent on the accumulation of mature BldA or is otherwise regulated by other BldA-dependent regulators.

All other genes with the exception of the five AHFCA regulatory cassette genes and the MFS transporter *sclE* are directly involved in SCL biosynthesis. As for *S. avermitilis*, the functional role and targets of the AHFCA receptor SclM1 have not been yet been identified, although its target AREs have been bioinformatically predicted (Chapter 3.3.1). Again, deletion of the pseudo-AHFCA receptor *sclM4* indicated it has an important role in regulating biosynthesis of SCL. No AHFCAs have thus far been observed in Δ*sclM4* mutants, so it is not known if AHFCAs are the autoregulators produced by this regulatory cassette, or that AHFCAs can induce the production of SCL.

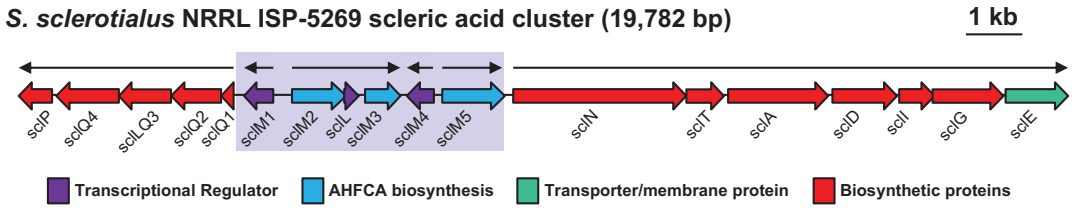


Figure 1-12 – Genetic organisation of the (putative) AHFCA-regulated scleric acid (*scl*) gene cluster in *S. sclerotialus* NRRL ISP-5269. The AHFCA regulatory cassette is enclosed in a blue box. Putative functions of *scl* genes are listed in Table 8-3, Appendix A.

1.3.5 *Streptomyces* ArpA-subfamily repressors coordinate complex regulatory networks

The *cpk* gene cluster (*SCO6264-6288;* Figure 1-13A) ArpA regulators ScbR and ScbR2 primarily control CPK biosynthesis through repression of *cpkO*, the gene cluster-specific transcriptional activator, which regulates the expression of the biosynthetic enzymes, including a three-gene type 1 polyketide synthase (*cpkABC*). Induction with SCB autoregulators initiates biosynthesis by binding to and releasing the GBL receptor ScbR from AREs located upstream of divergent *scbR/scbA-scbB*, within the coding frame of *scbA*, upstream of *cpkM* and at two sites upstream of *cpkO* (*cpkO$_A$ and cpkO$_B$*)[76, 98]. Within the gene cluster the pseudo-GBL receptor ScbR2 was initially reported to bind to AREs upstream of the *scbR/scbA* and *cpkO$_B$* promoters. Additional sites were later identified upstream of *cpkM*, *accA1*, *cpkD*, *cpkI*/*cpkJ*, and *cpkN*[76]. ScbR and ScbR2 recognises AREs similar to those of confirmed MmfR-type AREs, but the ScbR2 motif is much less well-defined than that of ScbR (Figure 1-13B). To reiterate, Δ*scbR* does not result in overproduction of either SCBs or CPK[98], while Δ*scbR2* results in constitutive expression of the CPK biosynthetic pathway[121],

indicating that both regulators play a role in the regulation of coelimycin biosynthesis. *scbR* is transcribed as early as 18 hrs during early growth phase, and ScbR2 after 30 hrs.

Both ScbR and ScbR2 also have pleiotropic roles in the regulation of genes in other key pathways, including regulation of the *cda*, *act* and *red* gene clusters, as well as other metabolism, stress response and morphological differentiation pathways[76]. Transcriptomics and corroborating gel shift assays have shown that both ScbR and ScbR2 bind the upstream regions of *gap1*, *afsK*, *cprA,* *narG3*, *tcmA*, *accA2*, *pyk2*, *SCO1403*/*cvnA4* and *SCO3868*/*soyB1*. ScbR-unique AREs have been identified upstream of *nagE2*, *cdaR*, and *SCO6323*/*6324*. ScbR2 has a larger number of unique target AREs, which includes *cprB*, *SCO3067*/*sig15*, *rpsL*, *sigR*, *sigF*, *SCO2529*/*leuA*, *valS*/*SCO2616*, *absA1*/*SCO3224*, *cdaPSI*/*SCO3229*, *SCO1571*/*ArgH*, *SCO1698*/*soxR*, *SCO3616*/*ask*, *cvnA1*/*SCO5545*, *fabG3*/*SCO1347*, *atrA*, *rpsD*, *SCO4503*, *SCO4677*, *SCO7623*, *actII-4*, *adpA* (not to be confused with *S. griseus adpA*, for which the homologue in *S. coelicolor* A3(2) is *bldH*), *murC*, *redD*, *redZ*, *serS*, and *cvnA12*. This indicates that in addition to coordinating *cpk* expression with other antibiotic pathways, these repressors also coordinate with GlcNAc transport, morphological differentiation, glycolysis, Pi metabolism and N metabolism, amongst other processes It is probable that there are also other sites within the genome that have not yet been identified due to cooperative repression of target genes with other repressors/regulators.

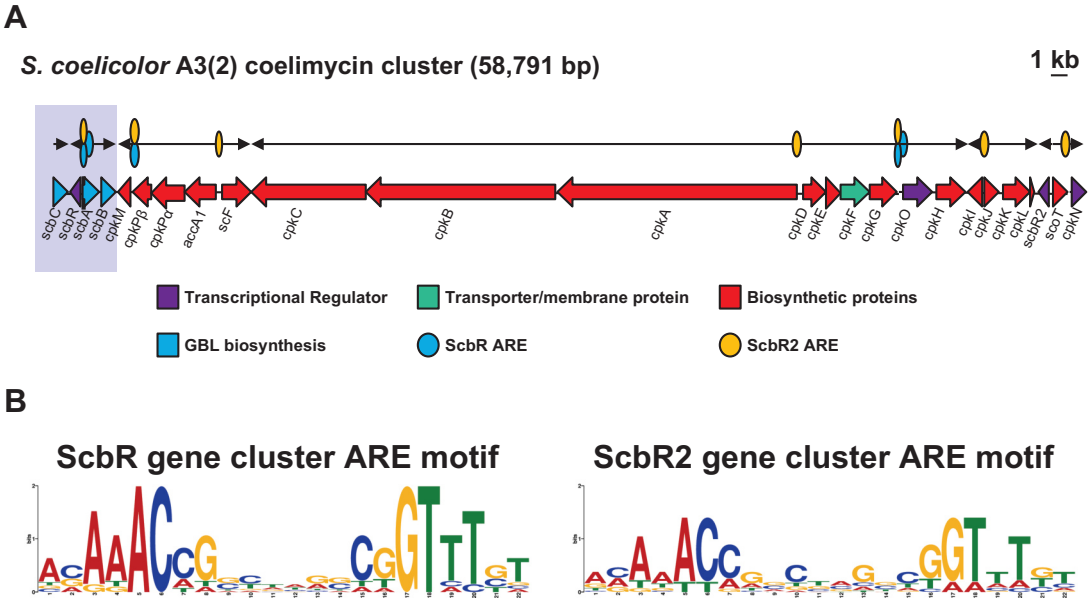


Figure 1-13 – Genetic organisation of the GBL-regulated coelimycin (*cpk*) gene cluster in *S. coelicolor* A3(2). (**A**) Annotated map of genes, operons and confirmed AREs bound by the GBL receptor ScbR and the pseudo-GBL receptor ScbR2. The GBL regulatory cassette is enclosed in a blue box; ScbR2 is not encoded within the regulatory cassette, but in the distal region of the *cpk* gene cluster. (**B**) MEME palindromic consensus motifs for ScbR and ScbR2 AREs bound within the gene cluster.

Just as ScbR2 regulates *cprB* transcription, CprB has been shown to bind AREs upstream of *cprB* and *cprA*, as well as *scbA* (within the ORF), $cpkO_A$ *and* $cpkO_B$ within the *cpk* gene cluster, presumably at the same ARE sites[69, 152]. Therefore, in addition to regulating *act*, CprB also regulates the *cpk* gene cluster. The experiments also showed that CprB can bind to an *S. griseus* ArpA consensus sequence. The CprB ARE recognition motif (Figure 1-14) is very similar to those of ScbR and ScbR2, which suggests that CprB may bind many of the same targets as either of these repressors. The proposed GBL effector for CprB has yet to be identified; despite being classed as a GBL receptor as a result of strong sequence homology with the LBD with the 6*R*-hydroxy-type GBL receptor ScbR, SCB2 was shown to be unable to dissociate the CprB:DNA complex[69]. However, organic extracts of cultured *S. coelicolor* A3(2) were able to disrupt the complex, suggesting an endogenously produced effector. Docking simulations performed by Bhukya *et al* suggest either a 6-keto-type GBL that is produced prior to reduction to 6*R*-hydroxy-type GBLs by ScbC, or it a novel GBL structure.

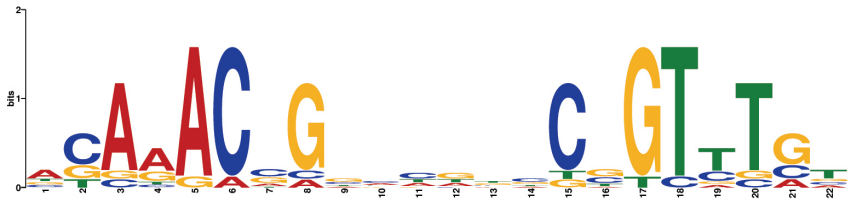


Figure 1-14 – MEME palindromic consensus for AREs bound by the ArpA-subfamily repressor CprB, a pleotropic regulator of secondary metabolite biosynthesis in *S. coelicolor* A3(2). The CprB motif is constructed from CprB binding to AREs regulating *cprB* (*OPB*), *cprA*, *scbA*, $cpkO_A$, $cpkO_B$ and an ArpA consensus sequence ($ArpA_{CS}$) bound during crystallisation studies[69].

In *S. avermitilis* the avermectin (*ave*) gene cluster (*saverm_933* to *saverm_953*) encodes for a type 1 polyketide synthase system, comprised of proteins AveA1 through AveA4, as well as additional tailoring proteins and the probable avermectin transporter AvtAB[153] (Figure 1-15A). The gene cluster itself encodes only a single regulator, the LuxR-family transcriptional regulator AveR, which functions as a transcriptional activator. The *ave* gene cluster was the first example of a biosynthetic gene cluster regulated by three ArpA-subfamily repressors working in tandem; AvaR1, AvaR2, AvaR3. As for *S. griseus*, the GBL regulatory cassette (*saverm_3702* to *saverm_3706*) is located far downstream of the *ave* gene cluster. Both Δ*avaR1* and Δ*avaR2* result in overproduction of both avenolide and avermectin[77, 78]. No other significant differentiation was observed for Δ*avaR1,* while Δ*avaR2* also resulted in stunted cell growth. Δ*avaR3* resulted in a reduction of the avermectin produced, but also an increase in filipin production[154]. Δ*avaR3* also disrupted cell growth and morphological differentiation.

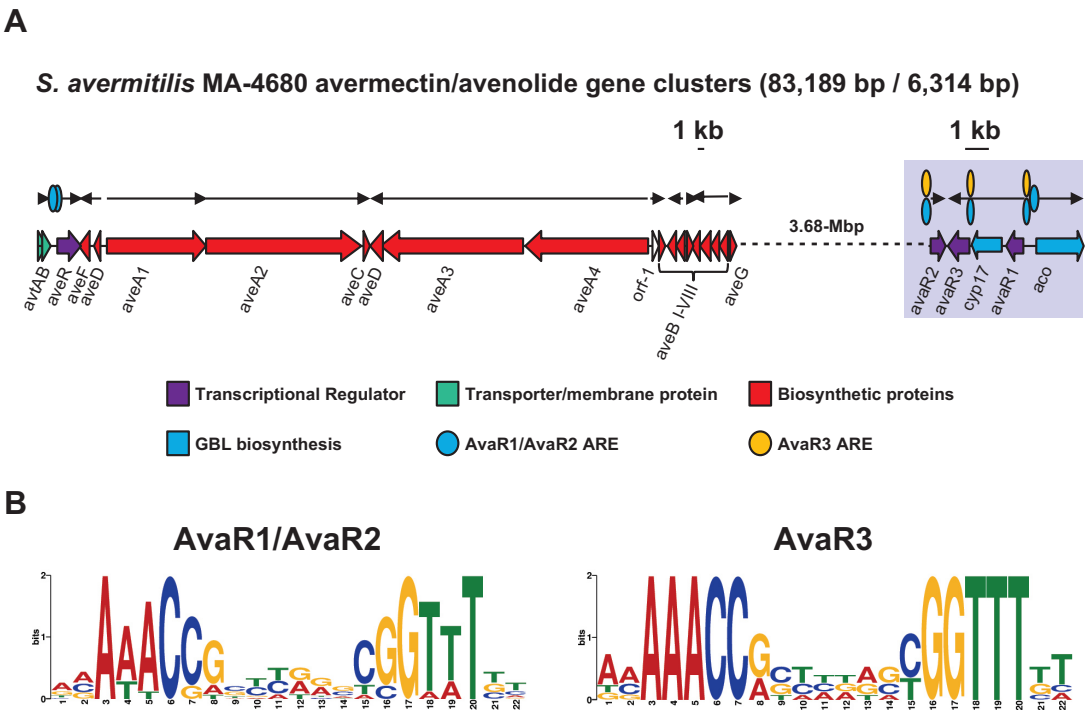


Figure 1-15 – Genetic organisation of the avenolide-regulated avermectin (*ave*)/avenolide gene clusters in *S. avermitilis* MA-4680. (**A**) Annotated map of genes, operons and confirmed AREs bound by the avenolide-receptor AvaR1 and the pseudo-avenolide receptors AvaR2 and AvaR3. The avenolide regulatory cassette is enclosed in a blue box. (**B**) MEME consensus palindromic motifs for AvaR1/AvaR2 and AvaR3 ARE sites within the avermectin and avenolide gene clusters.

In vitro DNase I footprinting assays show that AvaR1 and AvaR2 both bind to the same sites upstream of *avaR2*, *avaR3*, *avaR1-cyp17*, *aco*, and two sites upstream of *aveR* (*aveR-I* and *aveR-II*)[77, 78]. The sites are again similar to other ArpA AREs (Figure 1-15B). In addition, while both AvaR1 and AvaR2 interact with the autoregulator avenolide, AvaR2 is also receptive to exogenous JdB, making AvaR2 unique amongst pseudo-GBL receptors in that it is also regulated by the autoregulator that regulates the primary GBL receptor AvaR1[77]. AvaR3 does not bind to either *aveR-I*, *aveR-II*, or *aco*, instead binding only upstream of *avaR1-cyp17*, *avaR2* and *avaR3* at the same ARE sites as AvaR1 and AvaR2[154]. The cognate effector for the AvaR3 repressor is not known; AvaR3 also contains an additional stretch of amino acids within the LBD which is atypical of other TFRs, which likely has an unknown effect on ligand recognition.

Like ScbR/ScbR2, AvaR1 and AvaR2 have also been shown to have pleiotropic regulatory activity[77, 78]. Both repressors bind AREs upstream of s*averm_2675 (rpmB1), saverm_2685 (leuD), saverm_3490 (sig29), saverm_3560, saverm_3619 (aveT), saverm_4026 (amfC), saverm_4672 (folP2)*, *saverm_4838 (nuoB1)*, and *saverm_4935 (rpsQ)*. The only ARE thus far unique to AvaR1 is *saverm_6215 (cpdB)*, while AvaR2 has been shown to uniquely bind the promoter sites of *saverm_1230*, *saverm_2051* and *saverm_4075 (pstB)*. AvaR3 has been

suggested also act as a pleiotropic regulator and has been shown to regulate the production of phthoxazolin A, but target AREs have not been identified[154, 155].

Both ScbR/ScbR2 and AvaR1/AvaR2 bind DNA not only as homodimers, but also heterodimeric complexes. The existence of the ScbR:ScbR2 heterodimer was demonstrated by coimmunoprecipitation and was shown to bind an ARE upstream of *SCO5158* which is not bound by either homodimeric ScbR or ScbR2[156]. GST-pulldown of AvaR1 and AvaR2 tagged separately with GST and His tags and analysed using anti-GST and anti-His antibodies in western blot have shown that both the protein can complex together as a heterodimer, which has been shown to bind the *aveR*-II ARE both competitively as homodimers and cooperatively as a heterodimer in gel shift assays[78]. This adds an additional layer of complexity to these already complex ArpA systems that work in tandem with one another.

1.3.6 The current model of regulation in AHFCA-regulated systems

S1 nuclease protection analysis of RNA transcripts isolated from liquid cultures of *S. coelicolor* J1506 (SCP1+, SCP2-*hisA1 uraA str1 pgl-1*)[67] incubated with an organic extract of then-unidentified AHFCA indicated that *mmfR* and *mmfL* are expressed first, then maintained at a basal level until the accumulation of BldA enables rapid translation of *mmfR*/*mmfL* transcripts at ~48 hrs. After ~48 hrs the first signs of MmA production are observed, as are the first transcripts of *mmyR* and *mmyB*. Over the next ~24 hrs a rapid increase in transcription of *mmyB* observed, followed by an increase in the expression of the MmA biosynthetic genes. After ~72 hrs the transcript level of *mmyR* is at its highest, correlating with a decrease in the transcription of *mmfR* and *mmfL*. Given the known binding sites of MmfR and assumed binding sites of MmyR, an *in situ*, the model of regulation has been proposed as thus:

1) Until the onset of aerial hyphae formation, the *mmy* gene cluster is repressed by MmfR, which is maintained at a basal intracellular level. If AHFCA is already present MmfR is released from its targets, but *mmfL* and *mmyB* are not translated because of insufficient accumulation of BldA.

2) Once BldA matures, *mmfLHP* and *mmyB* can be translated, but not transcribed due to repression by MmfR. Above a critical concentration of AHFCAs enough MmfR is released from DNA that high expression levels of the *mmfLHP* genes result in sufficient AHFCA production at a rate faster than the production of MmfR protein, temporarily sustaining derepression of the gene cluster across a population of cells.

3) As MmfR is increasingly inhibited by AHFCAs *mmyB* is proportionally expressed, activating transcription of the biosynthetic enzymes of the gene cluster as well as its own expression, resulting in the production of MmA. MmA then binds to MmyJ to enable expression of the transporter MmR, through which MmA is exported from the cells.

4) As MmyR is increasingly produced, it represses transcription of *mmfR*, *mmfLHP* and *mmyB* by binding (presumably) to the same sites as MmfR. Decreasing levels of MmfLHP proteins are no longer able to sustain sufficient AHFCA levels to continue inhibiting MmfR, which then begins rebinding to its target AREs. At sufficient levels MmyR either outcompetes MmfR for available sites such that AHFCAs are rendered non-inducive, and/or MmfR and MmyR both bind to AREs to fully repress the *mmy* gene cluster.

Thus, MmfR acts as the activatory sensor for MmA production in response to quorum sensing, while MmyR acts as the terminator of MmA production (Figure 1-16). This rationalises the observed differences between Δ*mmfR* and Δ*mmyR*; without MmfR the MmyR repressor simply represses the system, while the MmfLHP enzymes self-sustain the inhibition of MmfR in the absence of the MmyR repressor. Like ScbR2, it is possible that MmyR also binds in the operons of the biosynthetic genes to shut down activation by lingering MmyB, but there is no evidence of this thus far. It is also possible that there are other ways to trigger AHFCA biosynthesis other than by direct induction with AHFCA.

It is proposed that both AHFCA and pseudo-AHFCA receptors also function as pleiotropic regulators to coordinate other secondary pathways with the biosynthesis of their own AHFCA-regulated gene clusters. The AHFCA receptor targets within the AHFCA-regulated gene clusters have greater palindromacy than other ArpA counterparts, so it is not known whether MmfR-type repressors are more specific for their target AREs than ArpA-subfamily repressors belonging to other systems. Additionally, it is possible that the MmfR and MmyR form both homodimeric and heterodimeric complexes to bind cooperatively at shared target AREs or by switching between configurations to be able to recognise configuration dependent AREs. Another aspect that has not been resolved is the identification of the MmyR effector, and what its function may be. It is probable that like ScbR2 and other pseudo-GBL receptors, MmyR functions as an antibiotic receptor. In the case of an exogenous antibiotic this may trigger further production of MmA as a survival response, but in the case of an endogenous antibiotic this is most likely to coordinate with expression of other biosynthetic pathways.

Of the other AHFCA-regulated gene clusters described, the azoxy compound gene cluster of *S. avermitilis*, given the distribution of TTA codons, likely functions under a similar model, albeit with more complex components given the increased complexity of the gene cluster. *S. venezuelae* probably functions similarly, possibly with SVEN_4184 acting in an analogous role to MmyB. As no TTA codons are found in the *scl* gene cluster, it is difficult to speculate on how this system is regulated with respect to the life cycle of *S. sclerotialus*.

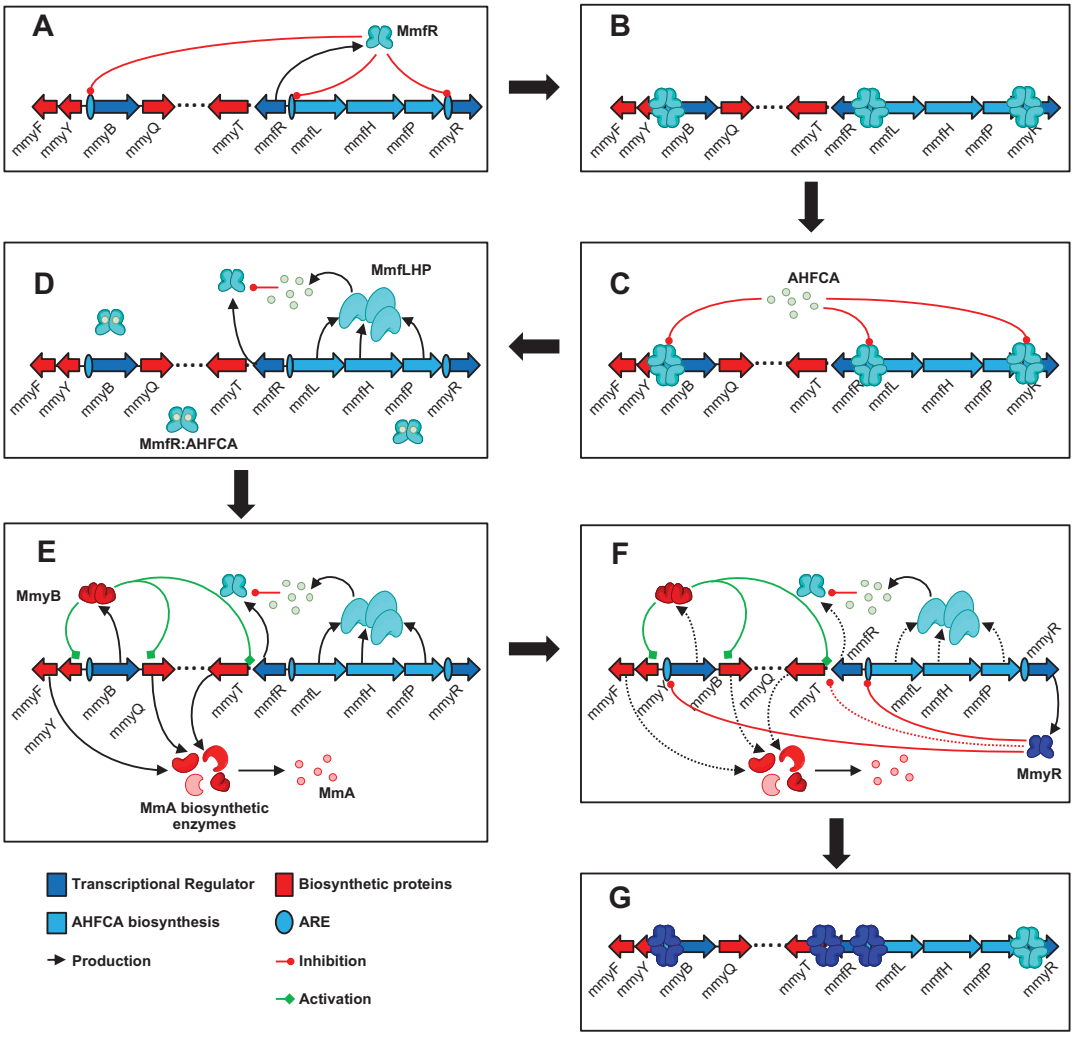


Figure 1-16 – Current model of *mmy* gene cluster regulation by MmfR/MmyR in *S. coelicolor* A3(2). (A and **B**) Expressed MmfR binds to AREs upstream of *mmfR*, *mmfLHP*, *mmyR* and *mmyB*, repressing transcription of the *mmy* gene cluster. Some transcription of *mmfL* is observed, but in the absence of BldA translation does not occur (as a result of a rare UUA codon found in the *mmfL* and *mmyB* mRNA transcripts). (**C**) In the presence of increasing cytoplasmic concentrations of AHFCA MmfR is released from DNA. Before accumulation of BldA *mmyB* and *mmfL* are increasingly transcribed but not translated. (**D**) At ~48 hrs matured BldA enables translation of the *mmfLHP* transcript. Expressed MmfLHP produces further AHFCAs that inhibit DNA-binding of expressed MmfR. (**E**) At the same time *mmyB* is translated, and the expressed MmyB activates transcription of the MmA biosynthetic enzymes. (**F** and **G**) Late expression of MmyR binds to the AREs upstream of *mmfR*, *mmfLHP* and *mmyB* (and/or other target sites), repressing transcription of all *mmy* genes.

1.4 Structural insight into the ArpA-subfamily repressors

1.4.1 Architecture of TetR-family transcriptional repressors

Figure 1-17 shows the crystal structures of the repressor TetR in complex with anhydrotetracycline and the *tetO* operator[157, 158]. Numerous crystallographic, oligomerisation and activity studies indicate that TFRs generally exist in solution as a combination of monomeric and dimeric states, the majority in the latter. Crystallography in particular shows that the dimeric configuration, often described as an "Ω" configuration, is the minimally 'active' state, as the rotation of these monomers is inherently required for both activation and deactivation of the TFR:DNA complex. TFRs, as mentioned prior are comprised of an N-terminal DBD and a C-terminal LBD, which also provides the main interface for dimerisation. There are nine canonical helices which form the core structure of all TFR structures; from N-terminus to C-terminus these are the stabilisation helix (α1), spacer helix (α2), recognition helix (α3), transmission helix (α4), ligand core helices 1, 2 and 3 (α5, α6 and α7), and dimerisation helices 1 and 2 (α8 and α9).

The DBD is comprised of a tetra-helical helix-turn-helix (HTH) motif formed by helices α1 to α4, with helices α2 and α3 forming the core HTH motif. The N-terminus of the α4 helix is also frequently involved in interactions with DNA. The HTH is a major structural motif found in many transcriptional regulators and is often found in proteins that also assemble in dimeric configurations[159]. The tetra-helical configuration as found in TFRs is also highly prevalent in prokaryotic regulators, such as cHTH, AraC, LuxR, TrpR and DnaA-family repressors[159].

The three ligand core helices (α5 to α7) are arranged in a triangular configuration forming the ligand-binding pocket (LBP) within the LBD. The two dimerisation helices (α8 and α9) run antiparallel along the inner-face of the exposed LBP, in a plane close to that of ligand core helix 1. Through these helices, and the C-terminal of ligand core helix 2, numerous electrostatic and Van der Waal interactions, and occasional covalent disulphide bridges form the main dimeric interface through which the dimer is formed around a two-fold rotational axis. On the distal side of the LBP is the transmission helix, which connects the DBD and LBD. The transmission helix is involved in key electrostatic interactions with the ligand; thus, the transmission helix is the only helix involved in both DNA and effector interactions.

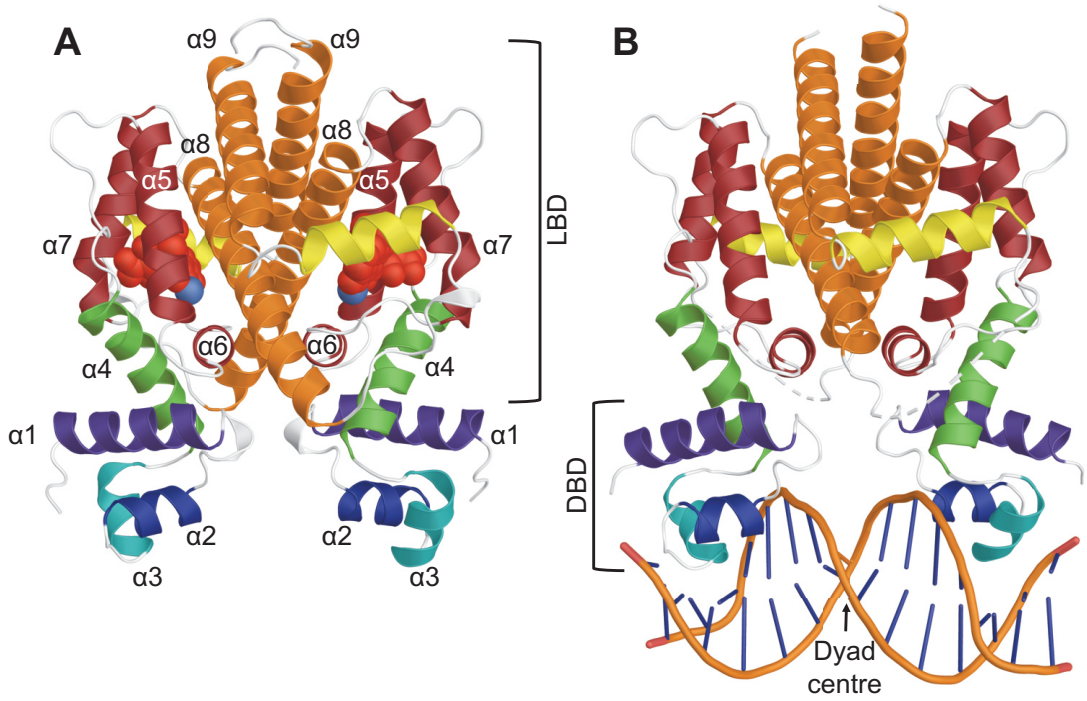


Figure 1-17 – Crystal structures of *Escherichia coli* TetR in complex with anhydrotetracycline/Mg^{2+} and *tetO*. (**A**) 1.9 Å resolution structure of homodimeric TetR in complex with anhydrotetracycline (**red sphere**) and Mg^{2+} (**blue sphere**) (PDB:4D7M), annotated with the nine canonical TFR helices; α1-α4 = helix-turn-helix (HTH) DNA-binding domain (DBD), α4-α9 = ligand-binding domain (LBD), α4 = transmission helix. The LBD is further split into α5-α7 = LBP core, and α8-α9 = dimerisation interface. (**B**) 2.5 Å resolution structure of homodimeric TetR bound to the *tetO* operator sequence (PDB:1QPI). TetR helices are coloured as **purple** (stabilisation helix), **blue** (spacer helix), **teal** (recognition helix), **green** (transmission helix), **red** (ligand core helices), **orange** (dimerisation helices), and **yellow** (non-conserved helices). The dyad centre for of the pseudo-palindromic *tetO* sequence is indicated with orange nucleotide bases.

1.4.2 DNA-binding mechanisms

The crystal structures of TFRs co-crystallised with target DNA sequences (Table 1-2) provide key insights into how TFRs recognise their target DNA sequences. Like other DNA-binding proteins, much fewer TFRs have been co-crystallised with target DNA sequences due to the inherent difficulty of crystallisation[160], a result largely of the flexible structures of the DNA molecules. Such co-crystallisations require extensive optimisation of oligonucleotide sequence length, palindromacy, protein:DNA molar ratios, solubility of the complex, *etc.* to acquire ordered crystals. TFRs can bind to DNA as either a single dimer or as a tetrameric assembly of monomers, more often referred to as a dimer of dimers. However, not all TFRs that bind as a single dimer can bind as a pair of dimers, and the binding of two dimers is thought to be necessary for TFRs which bind as pairs of dimers; thus, TFRs can be grouped based on binding stoichiometry.

Sequence length, palindromacy and composition all factor into DNA-recognition by TFRs. The minimal sequence length required for binding is usually a strong indicator of the valency of DNA-binding; single dimers recognise a mean sequence length of 14.3 bp ± 3.4 bp, while

a dimer of dimers recognises a mean sequence length of 21.3 bp ± 2.2 bp, based on range of electrostatic interactions formed. The palindromacy of the sequence is important as this orients the binding of TFR dimers symmetrically around the dyad centre. Sequence composition determines recognition through interactions with nucleotide backbones and bases *via* both electrostatic and Van der Waal interactions and is important in determining the rigidity ("bendability") of DNA.

Table 1-2 – Crystal structures of TFRs co-crystallised with their target operators available in the Protein Data Bank (PDB). PDB identifiers are indicated in parenthesis. Some TFRs, such *E. coli* SlmA are omitted due > 90 % sequence identity with other TFRs binding to the same sequences.

TFR (PDB identifier)	Regulatory Function	Dimer Pairs	Ref.
Staphylococcus aureus QacR (1JT0)	Multi-drug resistance	1	**(161)**
Escherichia coli TetR (1QPI)	Tetracycline resistance	1	**(157)**
Corynebacterium glutamicum CgmR (2YVH)	Multi-drug resistance	2	**(162)**
Pseudomonas aeruginosa PAO2 DesT (3LSP and 3LSR)	Membrane lipid homeostasis	1	**(163)**
Lactococcus lactis IL1403 HrtR (3VOK)	Heme homeostasis	1	**(164)**
Streptomyces antibioticus SimR (3ZQL)	Simocyclinone export	1	**(165)**
Vibrio cholerae O1 biovar El Tor N16961 SlmA (4GCT)	Cytokinesis	2	**(166)**
Thermotoga maritima ATCC 43589 TM1030 (4I6Z)	Putative Heat-shock	2	**(167)**
Mycobacterium smegmatis MC2 155 Ms6564 (4JL3)	Pleiotropic	2	**(168)**
Pseudomonas aeruginosa PAO5 PA2196 (4L62)	Putative Toxin resistance	2	**(169)**
Streptomyces coelicolor A3(2) CprB (4PXI and 5H58)	Pleiotropic	2	**(69, 152)**
Mycobacterium tuberculosis H37Rv KstR (5UA2)	Cholesterol catabolism	1	**Unpub.**
Myxococcus xanthus DK 1622 AibR (5K7Z)	IV-CoA biosynthesis	1	**(170)**
Bacillus halodurans C-125 FadR (5GPC)	Fatty acid metabolism	2	**(171)**

To facilitate orientation and anchoring to DNA the DBD surface often has a high distribution of positively charged amino acid residues, such as lysine or arginine. The recognition helix (α3), as its name suggests, has a pivotal role in DNA recognition. During binding the α3 helix docks into the DNA major groove to form specific interactions with nucleotide bases. Fitting into the groove is dependent on the local charge, geometry and torsional flexibility of the major grove to adapt to the space occupied by the α3 helix. The spacer helix (α2) is anchored to the phosphate backbone of the major groove through electrostatic interactions, facilitating the docking of the α3 helix by distortion and expansion of the major groove. The stabilisation helix (α1), as the name suggests, stabilises the DNA complex by formation of interactions with the phosphate backbone in distal regions of the TFR:DNA interface, and has little role in

DNA recognition. The α1 helix of some TFRs, such as HrtR, are not involved in direct interactions with DNA[164], while in others, such as CprB, deletion of the first six residues in the α1 helix disrupts DNA-binding[152]. Finally, the transmission helix (α4) may contribute interactions with the backbone through a highly conserved and promiscuous lysine residue. Of the two TFR domains the DBD is the most closely conserved across all TFRs; classification of TFRs into subfamilies is based on their DBD and the sequences they recognise, rather than the effectors to which they respond to.

Table 1-3 – Count of electrostatic interactions formed by TFRs when co-crystallised with target operator sequences. Interactions with DNA for each TFR were mapped manually in PyMOL with a cut-off of 3.6 Å (PDB identifiers as in Table 1-2). SR = stoichiometric ratio of dimers bound to DNA, ESI = electrostatic interactions, N_B = interactions with nucleotide backbones, N* = interactions with nucleotide bases, R_S = interactions *via* residue sidechains and R_B = interactions *via* residue backbones. Bases highlighted **blue** are the range of nucleotides for each strand interacting electrostatically with the TFR, with interactions *via* nucleotide bases in bold and underlined. Palindromacy of nucleotides around the dyad centre indicated in upper case. IUPAC nucleotides codes used in the table: k = G/T, m = A/C, s = G/C, r = A/G, w = A/T, and y = C/T.

TFR	Operator sequences	SR	ESI	N_B	N*	R_S N_B	R_S N*	R_B N_B	R_B N*
CprB (*ArpA$_{CS}$*)	5’-AcAtACGGGaC|GcCCCG**T**tTaT-3’ 3’-TgTaTG**C**CCt**G**|C**g**G**G**GCAaAtA-5’	2	33	30	**3**	20	3	10	0
CprB (*OPB*)	5’-AggcaggCgGcaC|GgtCt**G**ttgagtTc-3’ 3’-tccgtccGgCctG|CcaGaCaactcaAg-5’	2	27	26	**1**	15	1	11	0
TM1030	5’-GACT**G**ACTGACA|TGTCAGTCAGTC-3’ 3’-CTGACTGACTGT|ACAGTCA**G**TCAG-5’	2	30	28	**2**	16	0	12	2
Ms6564	5’-tCataAaCGAGACGg|t|aC**G**T**C**TCGT**c**TtgtG-3’ 3’-aGtatTtGCT**C**T**G**Cc|**a**|t**G**CAGAGCAg**A**acaC-5’	2	43	31	**8**	21	7	14	1
SlmA	5’-TTACGTGAGT|A**C**TCA**C**GTAA-3’ 3’-AATG**CA**CT**C**A|TGAGTGCATT-5’	2	36	31	**5**	22	5	9	0
FadR	5’-cAT**G**AATGAgT|AtTCATTCATc-3’ 3’-gtacttactcA|TaAGTAA**G**TAg-5’	2	39	37	**2**	22	0	15	2
QacR	5’-CTTATAGACCGATC|**G**ATCG**G**TCTATAAG-3’ 3’-GAATATCT**G**GCTAG|CTAGCCA**G**ATATTC-5’	2	48	44	**4**	32	3	12	1
PA2196	5 ’-g**t**ttcTA**G**ACg**A**|c|Tg**G**TCTAattca-3’ 3’-caaag**A**TCA**G**cT|g|Ac**C**AG**A**Ttaaga-5’	2	35	30	**5**	24	4	6	1
CgmR	5’-TA**A**CT**G**TACC**G**ACC|**GG**TCG**G**TACAGTTA-3’ 3’-ATTGACAT**G**GCTG**G**|CCA**G**CCAT**G**TCAAT-5’	2	36	27	**9**	16	9	11	0
DesT	5’-ktAcatc**AG**TgAACgc|ky**G**TTsACTsrwkTr-3’ 3’-maTgtagTCAcT**T**G**c**g|mrCAAsT**GA**sywmAy-5’	1	24	18	**6**	7	6	11	0
AibR	5’-CCTACCGA|TC**GG**TAGG-3’ 3’-GGAT**GG**CT|AGCCATCC-5’	1	20	16	**4**	10	4	6	0
KstR	5’-cCcAcTAGA**AC**g|tGTTCTAaTaGt-3’ 3’-gGgTgATCTTGc|a**CA**AGATtAtCa-5’	1	23	19	**4**	13	4	6	0
SimR	5’-TTCGTACG|s|C**GT**ACGAA-3’ 3’-AAGCA**TG**C|s|GCATGCTT-5’	1	22	18	**4**	12	4	6	0
TetR	5’-cCTATCA|a|TG**A**TAGa-3’ 3’-gGAT**A**GT|t|ACTATCt-5’	1	14	12	**2**	6	2	6	0
HrtR	5’-ATGACAC|t|GT**G**TCAT-3’ 3’-TACT**G**TG|a|CACAGTA-5’	1	14	12	**2**	8	2	4	0

The number of interactions formed directly with nucleotide bases varies greatly amongst TFRs (Table 1-3); some TFRs, such as CgmR and Ms6564, form many interactions with specific bases (9 and 8 respectively), while others, such as TetR, TM1030, HrtR and FadR, form fewer interactions (two interactions – one per monomer). The lowest number of interactions is formed by CprB in complex with the operator for CprB (*OPB*), which only forms a single base-specific interaction. However, when bound to an ArpA consensus sequence ($ArpA_{CS}$), the CprB repressor forms three specific base interactions with different nucleotides, demonstrating that the number and location of interactions formed vary depending on the sequence CprB binds to. In the case of TFRs like CprB factors other than specific nucleotide interactions must play some role in determining the DNA specificity of TFRs.

For all TFRs the number of specific base interactions is less than the total number of electrostatic interactions with the phosphate backbone; from as high as 25 % of all electrostatic interactions (CgmR, DesT) to as low as 3.7 % (CprB:*OPB*). In the absence of many specific nucleotide interactions in some TFRs electrostatic interactions with the backbone of DNA must do more than just provide binding stability. While there is a significant correlation between TFR binding stoichiometry and the total number of electrostatic interactions formed ($r(13) = 0.84$, $p < 0.01$), there is little correlation between the number of nucleotide base interactions and stoichiometry ($r(13) = 0.15$, $p = 0.60$) or between number of nucleic base interactions and number of electrostatic interactions ($r(13) = 0.41$, $p = 0.13$). TFRs that bind as a single dimer have generally higher specificity and affinity for target operators compared with TFRs binding as a dimer of dimers, which are proposed to act more regularly as pleiotropic regulators[172], a function driven by their more promiscuous specificity.

The prevailing model of TFR binding is the “clamp-and-click” model, where conformational twisting and unwinding of the helical architecture of the LBD dimer interface destabilises the dimer, allowing rotation of the monomers to improve the stability of the TFR:DNA complex; example of videos CprB undergoing this conformational shift during binding to the ArpA consensus sequence ($ArpA_{CS}$) are shown in the Electronic Supplementary Materials. The DNA curvature adapts with this twisting, resulting in unidirectional bending of the DNA strand around the dyad/dimer centre. This is corroborated in early experiments of TetR indicating it is the binding of TetR that causes the bending curvature of *tetO* [173]. It is not explicit whether the docking of the α3 helix of each monomer into the major grooves are a simultaneous or sequential process, but he prevailing hypothesis is the later. The sequential model of dimer binding suggests that the recognition and binding of DNA is observed as docking of the first TFR dimer α3 helix into the DNA major groove (clamping), followed by bending of DNA driven by the twisting of the protein dimeric interface to enable docking of the second α3 helix

into the DNA major groove (clicking). In both cases the primary purpose for twisting around the dimeric interface is to increase the stability of the complex by driving the intermediate complex over an energy barrier into a more minimal energy state, in which case a TFR with insufficient binding energy will not be able to stabilise its interaction with the DNA and will subsequently dissociate. Causes of insufficient energy include DNA rigidity, the angle-geometry insufficient, and the dimeric interface stability. Because of the energy input required to form the complex, this is often referred to as the tensed state, with the ligand-bound state referred to as the relaxed state. If the there is insufficient energy to drive the second α3 helix into the DNA major groove, and the TFR decouples from the DNA.

What determines TFR:DNA stoichiometry is not understood. While both types bind to palindromic sequences, only the binding of a single dimer is symmetrical around the dyad centre. While the arrangement of a pair of homodimers bound to DNA is also symmetrical around the dyad centre, the centre of binding of individual dimers is offset 2 bp to 3 bp each side of the dyad centre (Figure 1-18). Localised bidirectional bending is also induced by each dimer centred around the dimer interface, which results in a global unidirectional bending profile. TFRs bound as a single dimer often have greater DNA curvature than those found in bound with pairs of TFR dimers and may be bent either towards (*e.g.* DesT) or away (*e.g.* HrtR) from the protein dimer. TFRs bound as a dimer pairs often show less global distortion of DNA (*e.g.,* TM1030 or FadR), however DNA curvature can be observed in some structures (*e.g.,* CprB and Ms6564). Structures of TFRs with two dimers indicate that there is often a bending bias towards one dimer. Bending of the site by the first dimer binding may strain the DNA into a conformation that increases the affinity of the second dimer for DNA, the binding of which undergoes dimer twisting to relax the strain on the DNA. It may be the case that there is no universal system, and TFRs each have individual binding mechanisms. Unfortunately, there is a current lack in scientific methods and technologies to study this kind of DNA-TFR conformational dynamics in real time for fast kinetic reactions, yet alone visualise the specific molecular interactions between TFR and DNA.

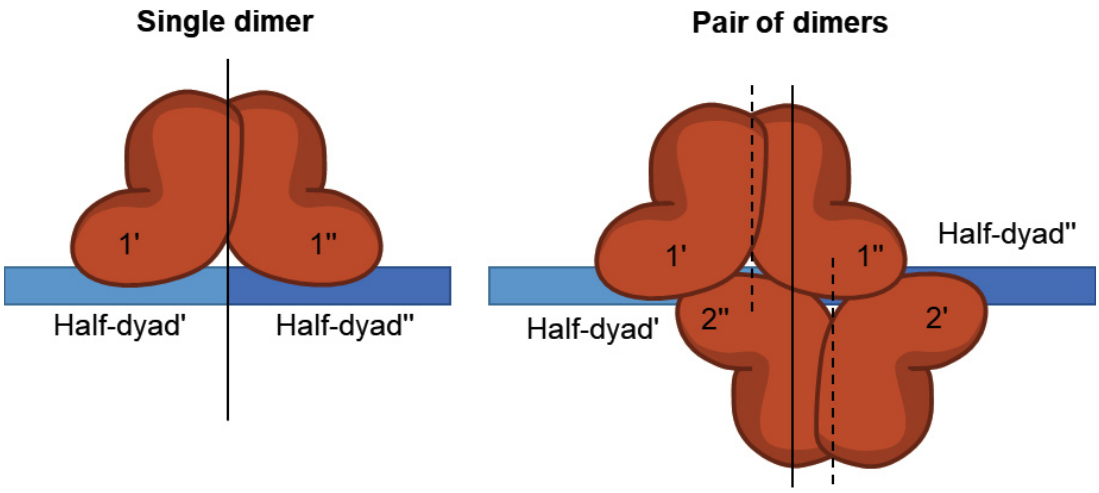


Figure 1-18 – Simplified model of binding symmetry for TFRs binding as either a single dimer or as a pair of dimers. Positions and lengths of TFR monomers are not to scale but demonstrate the concept. Each dimer is indicated by their number, with symmetry indicated by primes. The solid black line indicates overall symmetry of the complex, while the broken line indicates symmetry of each dimer.

1.4.3 Ligand-binding mechanisms

The layout of helices in the LBD results in an energetic relay through [dimer interface]-[LBP]-[α4]-[DBD]. Binding of the effector instigates dimer rearrangement through interactions with the residues of the dimerisation interface and neighbouring regions, as well as inducing a pendulum rearrangement of the DBD, with the α4 helix functioning as an energetic fulcrum between the LBD and DBD. This is made most apparent in comparison of crystal structures of TFRs in their native, DNA-bound and ligand-bound states. In TFR:ligand crystal structures this is observed as subtle changes in interdimeric rotation and distance of the DBD helices between monomers; often the ligand-bound structure has a conformation very similar to the native conformation, and will very frequently crystallise under similar conditions. The ligand inhibits TFR DNA-binding by stabilisation of the dimeric interface, incurring a conformational restriction that prevents the necessary bending of DNA as described in the clamp-and-click mechanism. Ligand binding also forces the release of the TFR from DNA, presumably through the same effects resulting in restabilisation of the dimeric interface and thus destabilises the interaction between the TFR DBD and DNA.

Poor homology of the LBD between TFRs results in large number of different effector specificities, broadening their potential roles as regulators in biological pathways. The helices are conserved in all these domains, however, some TFRs contain additional helices. TetR is one such example; an additional helix is found in TetR between dimerisation helices 1 and 2, which hooks around the LBP of the other monomer. This additional helix serves to stabilise the dimeric complex, as well as providing an additional interface for the expanded LBP which enables the binding of larger molecules, such as tetracycline and tetracycline analogues.

Because of the expanded pocket, short helical peptides are also able to bind in the LBP, which can mimic DNA-bound and effector-bound configurations[174].

While the inhibition of DNA-binding through effector binding is the most common TFR mechanism, alternative mechanisms have been observed. $_{rev}$TetR is a synthetic two-residue mutant of TetR that causes anhydrotetracycline to behave as a stimulator of DNA-binding, rather than as an inhibitor[175]. *Pseudomonas aeruginosa* DesT (PDB:3LSP) and *Myxococcus xanthus* AibR (PDB:5K7Z) also bind DNA only in the presence of their ligands[163, 170]. *Vibrio cholerae* SlmA is thus far unique in that DNA itself also behaves as the effector of SlmA; through the binding of DNA a non-canonical groove is formed on the surface of the LBD through which interactions with the C-terminus of the cytoskeletal cell division protein FtsZ (PDB:5HAW) are formed[176]. Although SlmA does have a small canonical LBP within the LBD, whether any ligand binds in this LBP is not known.

1.4.4 Crystal structure of *S. coelicolor* A3(2) MmfR

The 1.5 Å structure of the symmetrical homodimeric MmfR was solved by the Corre group in both its native and AHFCA2-bound conformations (Figure 1-19; Dr Rea, personal correspondence). The assigned helices are α1 (28-45), α2 (52-59) α3 (63-69), α4 (73-98) α5 (102-119), α6 (121-135), α7 (145-161), α8 (170-192) and α9 (195-211). For the AHFCA2-bound structure the assignments are the same except for helices α1 and α6, where unwinding of the n-terminus of helix α1 (30-45) and the c-terminal of helix α6 (121-132) is observed. In addition, a small coil is assigned in between helices α1 and α2 (47-49). Alignment of the MmfR dimers shows that the AHFCA2 bound conformation has a similar structure to the ligand-free conformation (1.5 Å RMSD over 1,485 atoms), with the greatest deviation in found in the DBD (2.7 Å RMSD over 346 atoms, compared with an RMSD of 0.9 Å (over 1139 atoms) for the LBD).

Like all TFRs, the DBD has a strongly conserved hydrophobic HTH sequence comprised of I34, L35, A38, F42, V52, V55, V66, and F70 (compared with CprB; I14, I15, A18, F22, L32, I35, L46 and F50). Regions contributing to the dimerisation interface are the α1-α2 loop (R48), α6 (R128, Q130, S131, E132), the α6-α7 loop (R133, A134, F135, L142, P143, Y144), α7 (V145), the α7-8 loop (R165, V168), α8 (A172, A173, R175, S176, A179, A180, F182, G183, M184, H186, V187, N190), and α9 (R200, L204, E203, E206, L207, M208, F210, A211, and L212). In its ligand-free form the dimeric interface has a surface area of 3,810 Å^2, and 3,794 Å^2 in the effector bound form. Hydrogen bonds are formed between the Y144 sidechain and

N190 sidechain, the R133 sidechain and N190 backbone, the R165 sidechain and E206 sidechain, and between the R175 sidechain and E203 sidechain.

The ligand-binding pocket of MmfR is comprised of residues H84 and Y85 in the α4 helix, L110, A113 and F117 in the α5 helix, M123, A127 and Q130 in the α6 helix, P143 and Y144 in the loop region between core helices 2 and 3 (helices α6 & α7), W147 in core helix 3 (helix α7), and V178, F181, F182 and Q185 in dimerisation helix 1 (helix α8). The solvent accessible surface area of the ligand-free MmfR LBP is 263 $Å^2$, and 239 $Å^2$ in the AHFCA2 bound form. For reference, AHFCA2 at pH 7.4 has a calculated Van der Waal surface area of 278 $Å^2$. The pocket entrance is an approximate elliptical hole with an area of 8.0 Å by 4.4 Å (~110.6 $Å^2$), which is partially obscured by the sidechain of H186.

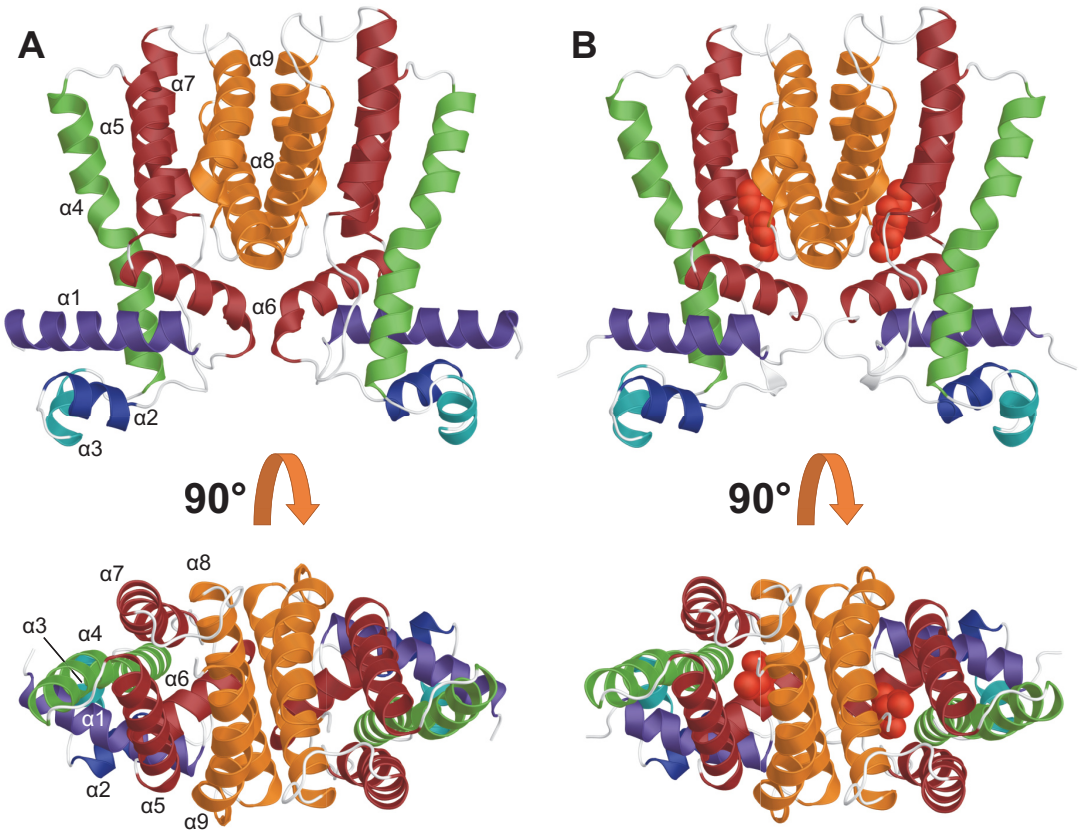


Figure 1-19 – 1.5 Å resolution crystal structure of homodimeric *S. coelicolor* A3(2) MmfR in unbound and AHFCA2-bound configurations. (**A**) Unbound and (**B**) AHFCA2-bounds, with colour coding as for Figure 1-17. Crystal structure was solved by Dr Rea (personal correspondence).

The structural differences between the two states is small compared to the change that would be expected for DNA binding. A conformational morph between MmfR in its native and AHFCA2-bound states is available in the Electronic Supplementary Materials. Between the ligand-free and AHFCA2-bound states there is a decrease in interchain helical centroid distances between helices α1 (-1.6 Å), α2 (-2.6 Å), α3 (-0.8 Å), α6 (-0.9Å) and α9, (-0.4 Å).

There is also an interchain increase in the angle between each helix, indicating subtle rotation of monomers around the interface, except for helices α3 and α4, which show a decrease of -9.2° and -0.8° respectively. Between adjacent helices within each chain there is an increase in angle for helices α1-α2 (4.1°), α5-α6 (4.6°) and α7-8 (1.7°), while a decrease is observed for helices α3-α4 (-7.1°) and α6-α7 (5.9°).

AHFCAs interact *via* hydrogen bonding of their 3-carboxylate moiety (calculated pKa 4.09) to two highly-conserved residues in the MmfR binding pocket; Y85 *via* its side-chain hydroxyl group and Y144 *via* its backbone amide (Figure 1-20). The hydroxymethyl group also interacts weakly with the Y85 side-chain hydroxyl group. Q130 stabilises the Y85-AHFCA interaction *via* hydrogen bonding of the side-chain amide (NH_2) to the side-chain hydroxyl group of Y85. Located nearby is H84, but its role, if any, has not yet been determined. HFCAs, with no alkyl chains, as well as other ArpA inducers such as SCB1, induce no activity *in vivo*, implying that the size and length of the alkyl chain, as well as the availability of a carboxylic acid group, is a determinant of ligand selectivity[149]. SCB1 has also been shown in gel shift assays not to release MmfR from DNA[148]. The Y85, Q130 and Y144 residues are conserved in the AHFCA receptor homologues SgnR, AvaL1 and SclM1, while these residues are not found in other GBL receptors, suggesting they are nonreceptive to AHFCA effectors. The specifics of how changes in the binding site affect AHFCA binding is further discussed in Chapter 6.

Y85A, Y85F, Q130E, Y144A and Y144F mutants have been investigated both *in vivo* and *in vitro* by Dr Styles and Dr Zhou respectively[148, 149]. *In vitro* both Y85A and Y85F mutants showed a significantly decreased response to AHFCA1 with similar affinities, indicating the necessity of the Y85 sidechain hydroxyl group. Q130E exhibited a similar decrease in activity, indicating the necessity of the amide group. Y144F had a similar/slightly increased ability to be released from DNA compared to the wild-type MmfR, while AHFCA binding was completely abolished in the Y144A mutant, presumably due to distorting of the LBP by substitution with a much smaller residue. These findings were corroborated by Dr Styles *in vivo*. Why Y144F improves AHFCA activity is not entirely clear. Dr Styles' data shows that Y144F has similar, if slightly weaker DNA-binding activity, even though the electrostatic interaction formed between Y144 and N190 of the other monomer is abolished in this mutant. The loss of this interaction may allow F144 to move into closer proximity of the LBP and thus form a stronger interaction with AHFCA molecules.

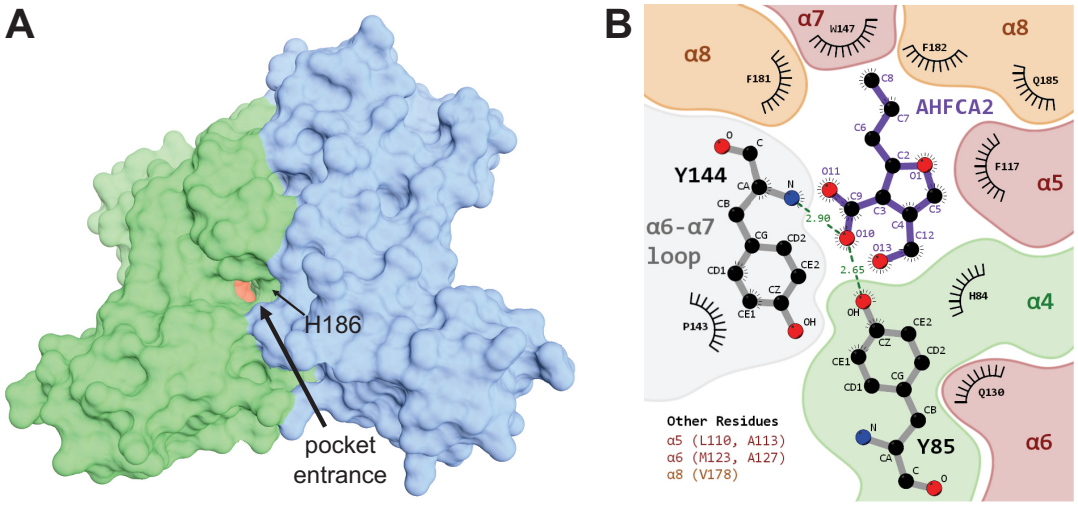


Figure 1-20 – Interactions of AHFCA2 with MmfR residues in the ligand-binding pocket (LBP) of the 1.5 Å resolution crystal structure of MmfR in complex with AHFCA2. (**A**) surface rendering showing the entrance to the LBP partially obscured by H186. (**B**) LigPlot+ rendering of the MmfR:AHFCA2 LBP. Some residues are not observed in LigPlot+ due to low proximity to the AHFCA2 ligand. Chemical valences not shown.

1.4.5 Crystal structures of GBL-receptors CprB and TylP

Of the TFR structures currently available in the Protein Data Bank (PDB), only two other ArpA-subfamily repressor structures have been solved thus far; *S. coelicolor* A3(2) CprB and *S. fradiae* TylP (Figure 1-21). CprB has been solved in both the ligand-free (1UI5 and 1UI6) and DNA-bound (PDB:4PXI and PDB:5H58) configurations[69, 152, 177], while TylP has been solved only in its ligand-free form (PDB:5XAY and PDB:5XAZ)[178]. TylP also influences morphological differentiation and is a repressor of tylosin biosynthesis in *S. fradiae*. Comparison of the native structures shows that despite belonging to the same TFR subfamily (35.2 % sequence identity with MmfR), the two repressors are very different in the arrangement of their helices. Compared with MmfR, the α1 helix of CprB is tilted downwards, while in TylP it is tilted upwards. In addition, both CprB and TylP have an additional α10 helix at the C-terminus that is not present in MmfR and other AHFCA receptors. TylP has a unique helix-loop extension that spans the length of the LBD down to the DBD; deletion of this helix results in abolition of DNA-binding[114]. The function of these extra helices is not currently known. The CprB dimer is held together by an atypical disulphide bridge between C159 of each monomer that is not to our knowledge found in any other ArpA regulator.

Despite having a homologous function MmfR is distinct from both CprB and TylP, although it more closely resembles TylP. RMSD using SUPER alignment in PyMOL is 3.6 Å over 2,044 atoms for TylP, compared with an RMSD of 4.3 Å over 1,868 atoms for CprB. This is corroborated in sequence homology, where MmfR shares 27.5 % overall sequence identity with CprB, compared with 31.8 % for TylP. The intradimeric distance between DBDs of the

MmfR dimer is an intermediate distance between CprB, which has a smaller distance, and TylP, which has a larger distance. The LBP of MmfR however, like CprB, is comprised of mostly hydrophobic residues, and MmfR, CprB and TylP all share a conserved tryptophan residue, whereas the TylP pockets is slightly more hydrophilic.

Like CprB, the cognate effector of TylP has yet to be identified. Attempted deregulation of the DNA-binding interaction between TylP and DNA with A-factor, SCB1, VB-C7 and IM-2-C7 all failed to release TylP from DNA, however organic extracts from wild type *S. fradiae* stimulated derepression, indicating an endogenous ligand[114]. Organic extracts of materials from a strain containing knockouts of *orf18** and *orf16** (encoding the Aco and Cyp17 homologues respectively) failed to release TylP from DNA, suggesting these two genes are responsible for producing the autoregulator, which is likely an avenolide-like GBL. Molecular dynamics have identified that a possible macrolide intermediate in the Tylosin biosynthetic pathways may also regulate the activity of TylP[178].

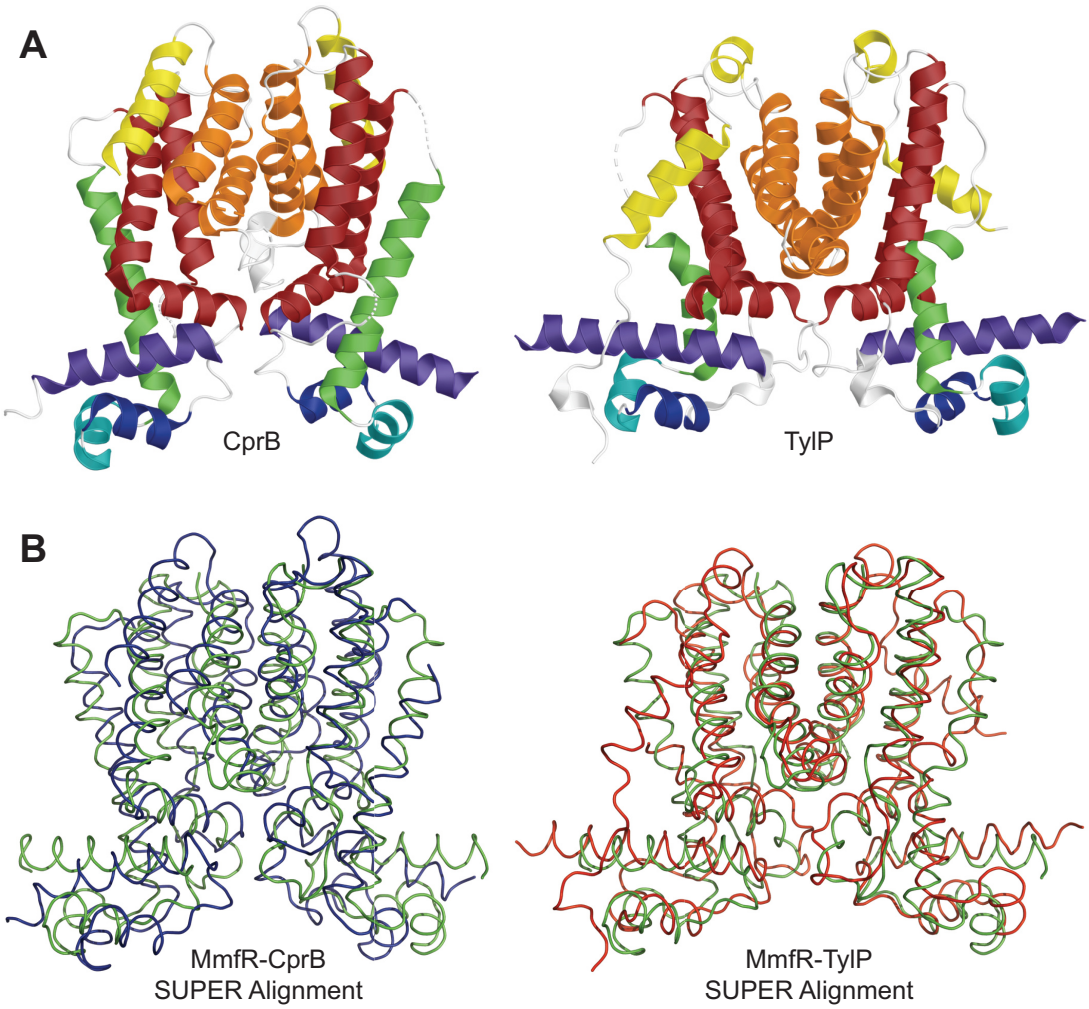


Figure 1-21 – Crystal structures for the unbound configurations of the ArpA-subfamily repressors *Streptomyces coelicolor* A3(2) CprB and *Streptomyces fradiae* TylP. (**A**) 2.4 Å resolution crystal structure of CprB and 2.3 Å resolution crystal structure of TylP. Colour coding as for Figure 1-17. (**B**) Ribbon structures of MmfR (**green**) SUPER aligned with CprB (**blue**; PDB:1UI5) and TylP (**red**; PDB:5XAZ) in PyMOL.

The crystal structure of CprB in complex with DNA (Figure 1-22) shows that it binds as a pair of dimers around a single ARE sequence, consistent for binding to either the *cprB* promoter or the ArpA consensus sequence motif. In CprB there is a bias in the global DNA bending towards one dimer, correlated by a difference in the degree of conformational shift of each homodimer between the native and DNA bound forms (RMSD of 2.9 Å and 4.0 Å over 384 Cα for each homodimer in DNA bound structure PDB:4PXI compared with native CprB structure PDB:1UI5). The structure also reveals that CprB binds asymmetrically to the sequence with respect to the dyad centre, the recognition helices having been shifted one base pair away from the centre though rotation around the major groove. The helices closer to the dyad centre are better docked to the major groove than those at the edge of the sequences, enabled by distorted expansion of the major grooving, possibly suggesting there is an order of binding for the two helices. The dimers do not lie flat on either side of the ARE but have a minimal angle of ~145° between dimers. Despite the observed difference in relative angles of the stabilisation helix, MmfR, CprB and TylP all bind similar recognition motifs. As mentioned before, deletion of six residues from the N-terminus of the α1 helix in CprB abolished DNA-binding, and therefore is necessary for binding. It is uncertain how necessary the N-terminus of the α1 helix is for either MmfR or TylP.

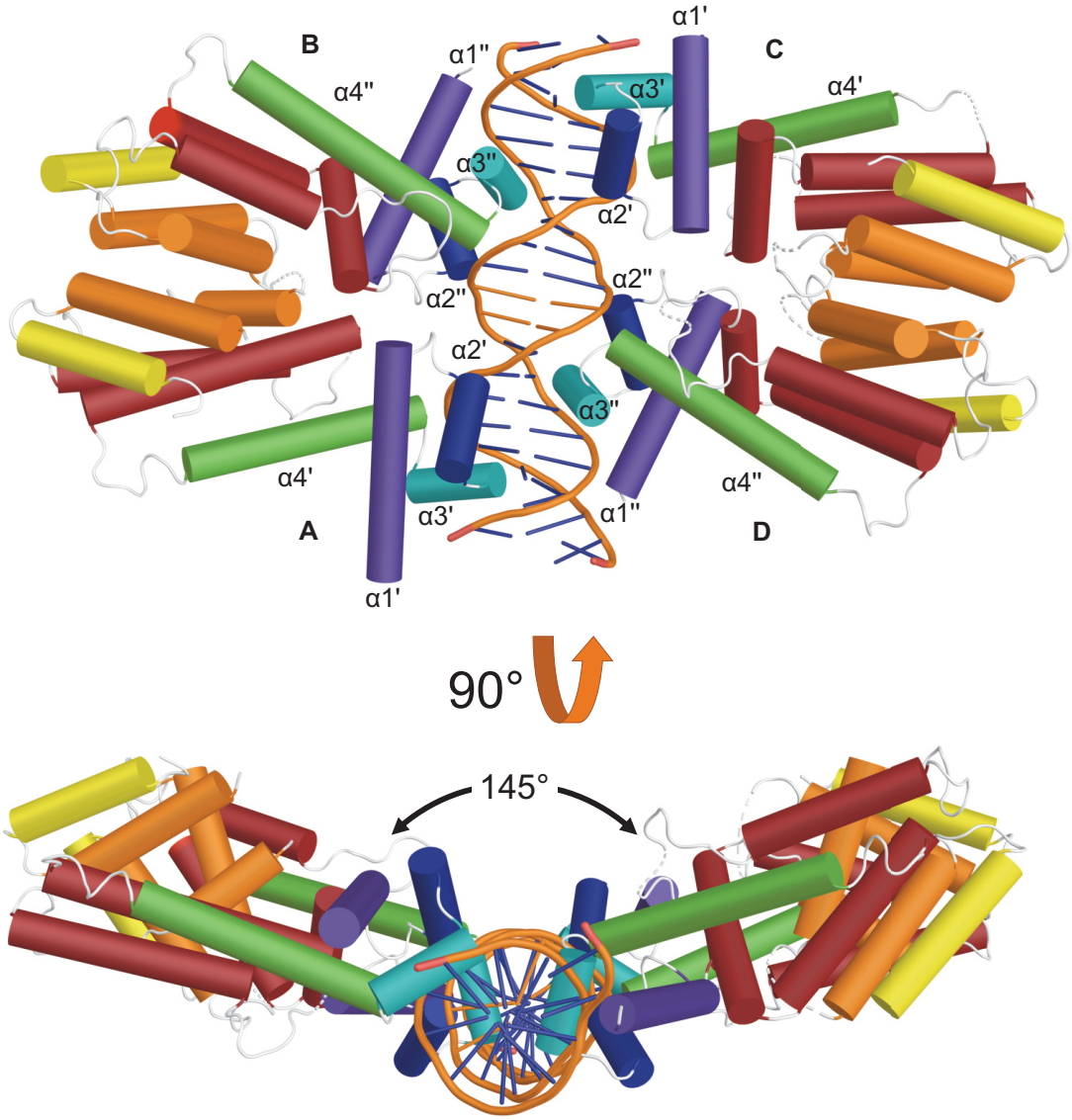


Figure 1-22 – 3.2 Å resolution crystal structure of *S. coelicolor* A3(2) CprB in complex with the ArpA consensus sequence (*ArpA*$_{CS}$). PDB identifier is 4PXI. Colour coding as for Figure 1-17. The dyad centre of the oligonucleotide is indicated by **orange** base ladders. Symmetry of CprB monomers is indicated by primes (').

1.5 Project Aims

1.5.1 Research questions & hypotheses

The Corre group is interested in elucidating the molecular mechanisms of AHFCA-regulated gene clusters to further our understanding of how these repressors function in relation to other pathways *in situ*, to discover new natural products and to develop novel biosynthetic systems exploiting the AHFCA regulatory cassette. The group is also interested in determining the similarities and differences in behaviour of MmfR and MmyR orthologues found in other actinobacteria species. For this PhD project, the goal was to characterise both the specific DNA and ligand-binding ability of these repressors *in vitro*, unimpeded by biological factors that complicate interpretation of their behaviour *in vivo*. To this end, the following questions were raised:

Deconvoluting regulation for AHFCA receptor / AHFCA receptor AREs

- What are the targets of receptor orthologues within *S. avermitilis* and *S. sclerotialus* AHFCA-regulated gene clusters?
- How do pseudo-AHFCA receptors control AHFCA-regulated gene clusters?
- Do these repressors act as pleiotropic regulators?

Hypotheses: It is anticipated that AHFCA receptors in other *Streptomyces* spp. will bind sites similar to that of MmfR, as was demonstrated *via* bioinformatic prediction of SgnR targets in *S. venezuelae*[148]. The AHFCA receptors in *S. avermitilis* (AvaL1) and *S. sclerotialus* (SclM1) are expected to bind to AREs within the AHFCA-regulated gene clusters. Similarly, it is expected that all AHFCA receptors are capable of binding targets cross-species, as has been observed previously for the *S. venezuelae* AHFCA receptor SgnR[148]. From *in vivo* experiments it remains unclear whether MmyR binds to the AREs sites as MmfR[149]; it is possible that the pseudo-AHFCA receptors from other *Streptomyces* spp. bind to the same targets as their partner AHFCA receptors. Both AHFCA and pseudo-AHFCA receptors are proposed to bind to some or all of the same AREs (with MmyR likely having reduced binding ability against these repressors in its native form. Like ScbR2[76], it is proposed that other targets of the pseudo-AHFCA receptor may be present within the AHFCA-regulated gene clusters. AHFCA and pseudo-AHFCA receptors, like their GBL/pseudo-GBL counterparts are also likely to be pleiotropic regulators[76-79]. These sequences are likely to be less alike to previously observed gene cluster-specific AREs.

Characterising ligand-binding in AHFCA receptors

- Are AvaL1 and SclM1 AHFCA receptors?
- Can AHFCA receptors recognise molecules other than AHFCA molecules?
- Does AHFCA receptivity change between AHFCA receptors of different species?
- How does the inhibitory activity of AHFCA molecules correlate with their structure?
- What are the cognate ligands of the pseudo-AHFCA receptors?

Hypotheses: Based on sequence homology it is proposed that MmfR, SgnR, AvaL1 and SclM1 all function as AHFCA receptors, with this functionality confirmed for the two former repressors[148]. It is expected that the AHFCA receptors are specific only for AHFCA molecules, and no other signalling molecules such as related GBLs. AvaL1, SclM1 and SgnR likely have an affinity for AHFCA effectors similar to that quantified *in vivo* for MmfR[149]. However, each AHFCA receptor may show a preference for AHFCAs produced by each *Streptomyces* spp.[95, 111, 112]. Any difference in apparent activity would be observed as a function of the 2-alkyl moiety of the AHFCA molecules. The pseudo-AHFCA are anticipated to function as antibiotic receptors, based on both sequence and functional homology with pseudo-GBL receptors acting as antibiotic receptors in GBL/pseudo-GBL receptor systems[97].

Investigating the DNA-binding mechanisms of AHFCA/pseudo-AHFCA receptors

- With what affinity do AHFCA/pseudo-AHFCA receptors bind target AREs?
- How specific is the recognition of target AREs?
- What is the binding stoichiometry of protein-DNA interactions?
- Through what mechanism is the protein-DNA complex formed?
- What components (amino acid specificity, electrostatics, geometry, *etc.*) of the DBD are key for DNA recognition?
- Are DNA-binding mechanisms for AHFCA/pseud-AHFCA receptors consistent with that of other ArpA-subfamily repressors?

Hypotheses: It is anticipated that both AHFCA and pseudo-AHFCA receptors have a DNA-binding affinity and mechanism consistent with that of other members of the ArpA-subfamily, for which *S. coelicolor* A3(2) CprB currently serves as the best model for quantitative analysis[69, 152]. DNA-binding is proposed to occur *via* a conformational change mechanism that is driven by both the twisting of the AHFCA/pseudo-AHFCA receptor homodimeric interface and *via* cooperative bending of DNA between two homodimers binding to the same target site. Given the possibility of being pleiotropic regulators, it is anticipated that the specify of the AHFCA receptors may be more degenerate than has previously been observed.

1.5.2 Outline of thesis

A simplified summary of the workflow is shown in Figure 1-23. Using polymerase chain reaction (PCR), the genes of AHFCA and pseudo-AHFCA receptors from *S. avermitilis* and *S. sclerotialus* were amplified and cloned into a pET151/D-TOPO® expression vector, and then expressed in an *E. coli* BL21(DE3) expression host. The recombinant proteins were then purified using immobilised metal affinity chromatography (IMAC), and gel shift assays used to assess their biological activity for both DNA-binding and inhibition of DNA-binding by AHFCA receptors. Quantitative characterisation, including derivation of affinity, kinetics, stoichiometry, thermodynamics and mechanical studies were investigated using surface plasmon resonance (SPR) on a Biacore T-200 machine. *In silico* simulations were performed to rationalise, corroborate and inform the interactions of these repressors with both DNA and ligands, based on the currently available crystal structure of MmfR, as well as new structures being generated through parallel crystallographic trials.

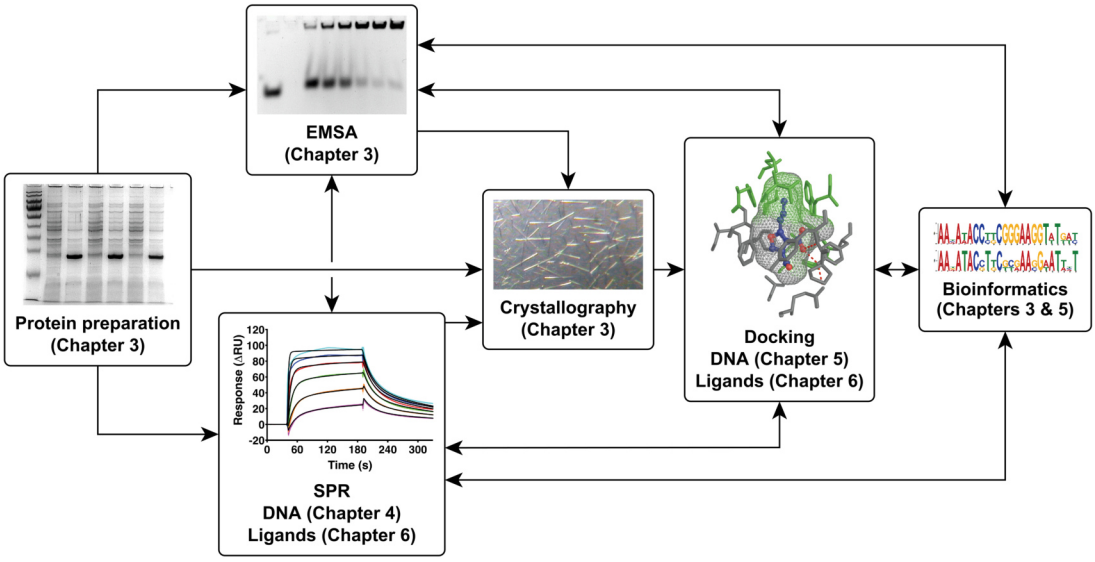


Figure 1-23 – General experimental workflow for biochemical investigation of MmfR and MmyR repressors. Results are described in the Chapters indicated in parentheses. Methodological details can be found in Chapter 2. EMSA = electrophoretic mobility shift assay, SPR = surface plasmon resonance.

Bioinformatic analysis was performed to predict AHFCA/pseudo-AHFCA receptor AREs in *S. avermitilis* and *S. sclerotialus*, based on known MmfR ARE sequences in *S. coelicolor* A3(2). From these predictions the sequences were tested and used to build a profile of the AHFCA/pseudo-AHFCA receptor binding motifs, which was compared with those of other ArpA-subfamily repressors. To investigate their roles as pleiotropic regulators, possible cross-talking with other ArpA-subfamily repressors, as well as identity other possible AREs within the AHFCA-regulated gene clusters, bioinformatics was again used to predict sites throughout the entire genome, and a methodology was designed using SPR to test binding.

Chapter 2 comprises the Materials and Methods for all experiments and procedures which are performed in subsequent chapters. Included are theoretical principles of how protein-DNA and protein-ligand interactions were investigated (EMSA and SPR), as well as principles of some of the applied statistical methods for data analysis. Chapter 3 focuses on the generation of soluble recombinant AvaL1, AvaL2 and SclM1 proteins for use in all *in vitro* biochemical assays, and subsequent functional tests using gel shift assays. The preliminary crystallisation trials for these repressors are also reported here.

Chapter 4 details the quantification the interactions between AvaL1/AvaL2 and target ARE sequences (identified in Chapter 3) using SPR. The affinity, kinetics, stoichiometry, thermodynamic stability and mechanism of AvaL1 binding to AREs were investigated using a streptavidin (SA)-chip-based methodology. Chapter 5 is a bioinformatics-led chapter focusing on the sequence characterisation of all confirmed AREs bound by AHFCA receptors, contextualised against AREs bound by other ArpA-subfamily repressors. Included is the prediction of other ARE targets within the genomes of *S. avermitilis* and *S. sclerotialus*. This chapter finishes with an *in silico* study of the molecular mechanisms of MmfR-DNA recognition using the crystal structures of CprB in complex with DNA as a reference.

Chapter 6 details the interactions of the AHFCA signalling molecules with the AvaL1 repressor using an SA-chip-based SPR method. Using other experimental data collected for effector activity against MmfR as well as *in silico* modelling of AHFCAs binding to MmfR and other AHFCA receptor orthologues, differences in MmfR and AvaL1 effector binding were characterised. This collective data was used to build a structure-activity relationship (SAR) profile for AHFCA signalling molecules binding to their target receptors.

Finally, Chapter 7 summarises the findings of this work and places them in context of the current work and literature surround AHFCA-regulated and other ArpA-subfamily repressor systems. Possible future work projects are also discussed.

2 Materials and Methods

2.1 Materials, equipment and software

Table 2-1 – Services, consumables and kits used in this project.

Supplier	Material
Bio S&T	*Streptomyces avermitilis* pESAC13A-19K fragment
Biotium	GelRed® Nucleic Acid Gel Stain
Day-Impex Ltd	Virkon® Disinfectant
Expedeon	InstantBlue™ Protein Stain
GE Healthcare	Biacore maintenance kit
	Ni-Sepharose™ High Performance
	Series S Sensor Chip NTA
	Series S Sensor Chip SA
Huntsman Corporation	Jeffamine SD-2001 Polyetheramine (100 % Solution)
Invitrogen	One Shot® TOP10 Chemical Transformation kit
	Champion™ pET Directional TOPO® Expression Kits
Jena Bioscience	JBS Deep Purple
Merck-Millipore	Amicon Ultra-15 Centrifugal Filter Unit MWCO 10 kDa
	Amicon Ultra-0.5 Centrifugal Filter Unit MWCO 10 kDa
	0.22 µm Syringe Filter
Molecular Dimensions	24-well XRL Plate
	Polystyrene MRC 2-drop Plates
	MIDAS™ (MD1-59)
	Morpheus® (MD1-46)
	JCSG-*plus*™ (MD1-37)
	PACT *premier*™ (MD1-29)
	Wizard Classic 1 and 2 (MD15-W12-T)
MP	FastDNA™ SPIN Kit for Soil
New England Biolabs® Inc.	Phusion® High-Fidelity DNA Polymerase
Sigma Aldrich	Custom DNA Oligo Tubes
Thermo Fisher Scientific™	Acrylamide:Bis-Acrylamide 29:1 (100 % solution)
	FastRuler™ DNA ladders LR and MR
	GeneArt™ Strings™ DNA Fragments
	GeneJET™ Plasmid Miniprep kit
	GeneJET™ Gel extraction kit
	PageRuler™ Prestained 10-180kDa Protein Ladder
	Taq DNA Polymerase (recombinant)
	10X *Taq* buffer + $(NH_4)_2SO_4$
	6X DNA Loading Dye

Table 2-2 – Equipment used in this project.

Supplier	Equipment
Beckman Coulter	Avanti J-25 Centrifuge
Bio-Rad Laboratories	Econo-Column® 1.0 x 30 cm Chromatography Column
Ltd	Mini-PROTEAN® Tetra Cell and Handcast System
	Wide Mini-Sub Cell GT Cell
Cole-Parmer	StableTemp Water bath, 2 L
Eppendorf	5424 R Microcentrifuge
	Mastercycler® Nexus Thermocycler
	Mastercycler® epGradient Thermocycler
Fisher Scientific	Fisherbrand™ accumet™ AB15 Basic and BioBasic™ pH/mV/°C Meter
	NanoDrop™ 2000 Spectrophotometer
GE Healthcare	Biacore T200
Grant	JB1 Water bath
INFORS HT	Multitron Incubator Shaker
New Brunswick™	C24 Incubator Shaker
	Innova® 44/44 Incubator Shaker
TTP Labtech	Mosquito® LCP

Table 2-3 – Software used in this project.

Developer	Software Package	Ref.
Actelion Pharmaceuticals Ltd.	DataWarrior v04.07.02	(179)
Bonvin Lab, Computational Structural Biology Group, Utrecht University	3D-DART	(180)
	CPORT	(181)
	HADDOCK	(182)
ChemAxon	MarvinSketch	(183)
GE Healthcare	Biacore T200 Control and Evaluation Software 2.0	
GraphPad Software	GraphPad Prism v8	(184)
GSL Biotech LLC	SnapGene® Viewer v4.1	(185)
Microsoft	Office 365 ProPlus	
Molecular Graphics Laboratory, The Scripps Research Institute	AutoDock v4.2.6	(186)
	AutoDockTools v1.5.6	(187)
	AutoDock Vina v1.1.2	(188)
National Centre for Research Resources	MEME Suite v5.0.0	(189-192)
Open-Source	ImageJ v1.52c	
Pennsylvania State University	MEGA X	(193)
PerkinElmer Informatics	ChemDraw Professional v16.0.1.4	
Schrödinger, LLC	Open-Source PyMOL v1.9.x (Win)	(194)
	MacPyMOL X11 Hybrid (Mac)	
Structural Bioinformatics Group, Imperial College London	Phyre^2	(195)
Thomson Reuters	EndNote® Cite While You Write v18.0.0.10063	
TTP Labtech	Mosquito® Control Software	
Department of Genome Sciences/Biology, University of Washington	PHYLIP v3.695	(196)
Swiss Institute of Bioinformatics	SwissDock	(197, 198)

2.2 Bacterial strains, constructs and sequences

Table 2-4 – Organisms and strains used in this project.

Organism	Genotype
Streptomyces avermitilis MA-4680	Contains the native plasmid SAP1
Streptomyces sclerotialus NRRL ISP-5269	n/d
Escherichia coli TOP10	F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) Φ80*lacZ*Δ*M15* Δ*lacX74 recA1 araD139* Δ(*araleu*)7697 *galU galK rpsL* (StrR) *endA1 nupG*
Escherichia coli BL21(DE3)	F- *ompT hsdSB* (*rBmB*-) *gal dcm* (*DE3*)

Table 2-5 – **Primers used in the amplification of MmfR/MmyR orthologues from *S. avermitilis* and *S. sclerotialus*.** For each primer their sequence orientation is indicated (F for forward and R for reverse). T_M values for annealing to the intended target were predicted using the SnapGene software. Primers in bold are those used for final amplification of target genes. Nucleotides in **bold** are altered starts/ends not present in the template sequence, while nucleotides underlined are undesired 5'-CACC-3' sequences. The number of sites is indicated for primer matches with at least a 8-bp match at the 3' end; primers for *ava* genes are for sites in the *S. avermitilis* linear chromosome, primers for *scl* genes are for sites in all available contigs, and for the *F243*$_F$/*EUB518*$_R$ primers sites are indicated as for *S. avermitilis*/*S. sclerotialus* based on previously stated criteria. Only a single site is observed for primers targeting codon-optimised (co) genes are based on the synthetic oligonucleotide template. T7 primers are used for confirmation of insertion into the pET151 vector.

Primer	Sequence (5' to 3')	Annealing Length	% GC	T_M (°C)	Sites
F243$_F$	GGATGAGCCCGCGGCCTA	18	72	63	7/53
EUB518$_R$	ATTACCGCGGCTGCTGG	17	65	61	173/1697
***avaL1*$_F$**	**CACC**ATGGACGTCATGAGTAGC	18	55	54	1
avaL1$_{FD3}$	**CACC**ATGAGTAGCGAGCGCAAC	19	59	59	5
avaL1$_{FL}$	**CACC**ATGGACGTCATGAGTAGCGAG	21	56	59	1
***avaL1*$_R$**	TCAGTGCAGTGGCACGAAGGC	21	62	63	2
avaL1$_{RD1}$	**TCA**CAGTGGCACGAAGGCG	16	63	60	14
avaL1$_{RS}$	TCAGTGCAGTGGCACGAAG	19	58	59	1
avaL2$_F$	**CACC**GTGGAGCCGGCAAAGGAG	20	68	65	2
***avaL2*$_{FS}$**	**CACC**GTGGAGCCGGCAAAG	17	68	61	1
avaL2$_R$	TCAGCGGGCCGGGGCGCCG	19	89	73	1,300
avaL2$_{RD7}$	**TCA**CGCCTGAGCCAGGGC	15	72	61	20
***avaL2*$_{RS}$**	TCAGCGGGCCGGGGC	15	87	64	696
sclM1$_F$	**CACC**AAGCAGGAACGCGCACGG	18	68	64	1,834
sclM1$_{FS}$	**CACC**AAGCAGGAACGCGCACG	17	67	62	2,026
sclM1$_R$	TCAGGGCATGGCGCCGC	17	76	65	4,641
sclM1$_{R2}$	TCAGGGCATGGCGCC	15	73	59	471
sclM4$_F$	**CACC**ATGGTCAAACAGGACAGG	18	55	54	79
sclM4$_{FL60C}$	**CACC**ATGGTCAAACAGGACAGGGC	20	58	59	257
sclM4$_{FL70C}$	**CACC**ATGGTCAAACAGGACAGGGCGCGACG	26	63	69	2,020
sclM4$_R$	TCACCGCACGGGGCGCCG	18	83	70	4,229
sclM4$_{RD1}$	**TCA**CACGGGGCGCCGCATC	16	74	65	1,061
sclM4$_{RS}$	TCACCGCACGGGGC	14	79	58	670
***sclM1*$_{CO-F}$**	CACCATGGATAGCAAAGCAGCAACC	25	52	62	1
***sclM1*$_{CO-R}$**	TTACGGCATTGCTCCGCGAAAAACT	25	61	64	1
***sclM4*$_{CO-F}$**	CACCATGGTTAAACAGGATCGTGCA	25	48	59	1
***sclM4*$_{CO-R}$**	TTAACGAACCGGACGACGCATCAGA	25	52	63	1
***T7*$_F$**	TAATACGACTCACTATAGGG	20	40	49	1
***T7*$_R$**	TAGTTATTGCTCAGCGGTGG	20	50	56	1

Table 2-6 – Plasmid vectors used in this Project.

Vector	Extra details	Size (bp)	Ref.
pET151	**Induction system:** T7*lac* promoter (*lacO*) **Inducer:** IPTG **N-Fusion Tag:** His_6 Tag, V5 Epitope Tag, TEV Protease Cleavage Site, Directional TOPO® Cloning Site **Also contains:** *amp* resistance marker, *lacI*, *rop*, *pBR322 ori*	5,760	(199)
pET151-*avaL1*		6,417	This work
pET151-*avaL2*		6,381	This work
pET151-*sclM1*$_{CO}$	*E. coli* codon-optimised *sclM1* sequence.	6,393	This work
pET151-*sclM4*$_{CO-1}$	*E. coli* codon-optimised *sclM4* sequence. Contains a 22-bp insertion resulting in R195 (non-recombinant) being substituted with QGRAQIRLLTKPERKLSWLLPPLSNN.	6,376	This work
pET151-*sclM4*$_{CO-3}$	*E. coli* codon-optimised *sclM4* sequence. Contains a single nucleotide deletion resulting in substitution of the stop codon with KGELRSGC extension at C-terminal.	6,353	This work
pET151-*sgnR*	Provided by Dr Zhou.	6,303	(148)
pET151-*mmfR*$_{Y85F}$	Provided by Dr Zhou.	6,405	(148)
pET151-*mmfR*$_{Y85A}$	Provided by Dr Zhou.	6,405	(148)
pET151-*mmfR*$_{Q130E}$	Provided by Dr Zhou.	6,405	(148)
pET151-*mmfR*$_{Y144A}$	Provided by Dr Zhou.	6,405	(148)
pET151-*mmfR*$_{Y144F}$	Provided by Dr Zhou.	6,405	(148)

Table 2-7 – Self-annealing oligonucleotides used in EMSA experiments. Nucleotide colours are indicated for the pyrimidine spacer (**orange**; serving as an attachment point for biotin), the GC clamp (**blue**) and the hairpin loop (**purple**). Nucleotides palindromic (**green**) non-palindromic (**red**) around the dyad centre (**black**) are also indicated.

Oligonucleotide	Sequence (5' to 3')
saverm_2301$_{ARE}$	CTCTCTGCAAAATACCTTCTCAAAGGAATTATGCGAGGCATAATTCCTTTGAGAAGGTATTTTGC
avaL1/avaA$_{ARE}$	CTCTCTGCAATATACCTGCGCGAAGGTATATTGCGAGGCAATATACCTTCGCGCAGGTATATTGC
avaL2/avaBC$_{ARE}$	TCTCTCGCAAGATACGTTCGTGAACGAATTCTGCGAGGCAGAATTCGTTCACGAACGTATCTTGC
SAV_CS_V1$_{ARE}$	GCAAAATACCTTCGCGAAGGAATTTTGCGAGGCAAAATTCCTTCGCGAAGGTATTTTGC
SAV_Pal_V1$_{ARE}$	GCAAAATACCTTCGCGAAGGTATTTTGCGAGGCAAAATACCTTCGCGAAGGTATTTTGC
sclM1/sclM2L3$_{ARE}$	TCTCCTGCAATATACCTATGGGAAGGTATATTGCGAGGCAATATACCTTCCCATAGGTATATTGC
sclM4/sclM5$_{ARE}$	CTCTCTGCAAGAAACTTCCAAGGCGGATCTTTGCGAGGCAAAGATCCGCCTTGGAAGTTTCTTGC
sclQ1-4P$_{ARE}$	TCTCTCGCTTGAAACCTTCGCGAAGGTTTGATGCGAGGCATCAAACCTTCGCGAAGGTTTCAAGC
sclN-E$_{ARE}$	CTCTCTGCTGAATACCTCCGCGAAGGAATGATGCGAGGCATCATTCCTTCGCGGAGGTATTCAGC

2.3 Recipes

2.3.1 DNA gel electrophoresis

1 % Agarose Gel Mixture
1 % w/v Agarose
Make up to volume in 1X TAE Buffer

GC Rich Sample Mixture
10 µL of 5X Phusion® GC Buffer
1 µL of 10 mM dNTPs
1 µL of 10 µM Forward Primer
1 µL of 10 µM Reverse Primer
1 µL of DNA template (variable concentration)
1 µL of Phusion® HF Polymerase
2.5 µL of DMSO
Make up to 50 µL with dH_2O

50X TAE Buffer
2 M Tris Base
1 M Acetic Acid
50 mM EDTA
pH to 8.3

pET151/*lacZ* Cloning Control Mixture
5 µL of 10X Taq buffer + $(NH_4)_2SO_4$
3 µL of 50 mM $MgCl_2$
1 µL of 10 mM dNTPs
0.5 µL of 0.1 µg/µl Control PCR Template
0.5 µL of 0.1 µg/µl Control PCR Primers
1 µL of *Taq* DNA Polymerase
Make up to 50 µL with dH_2O

Mixtures stored at 4 °C

2.3.2 Cell media and lysis buffers

LB Media
1 % w/v Tryptone
0.5 % w/v Yeast Extract
1 % w/v NaCl

SOC Media
2 % w/v Tryptone
0.5 % w/v Yeast Extract
10 mM NaCl
2.5 mM KCl
10 mM $MgCl_2$
10 mM $MgSO_4$
20 mM Glucose

LBA-AMP
1 % w/v Tryptone
0.5 % w/v Yeast Extract
1 % w/v NaCl
1.5 % w/v Agar
0.1 % w/v Ampicillin

Cell Lysis Buffer
50 mM KH_2PO_4/K_2HPO_4
400 mM NaCl
100 mM KCl
10 % v/v Glycerol
0.5 % v/v Triton X-100
10 mM Imidazole
pH to 7.8

Buffer stored at 4 °C.

2.3.3 SDS-PAGE buffers and gel mixtures

2X SDS-PAGE Loading Buffer
125 mM Tris-HCl
20 % v/v Glycerol
4 % v/v β-Mercaptoethanol
3 mM Bromophenol Blue
4 % v/v SDS
pH to 6.8

Running buffer stored at 4 °C. SDS-PAGE Loading Buffer stored at -20 °C.

10X SDS-PAGE Running Buffer
250 mM Tris-HCl
2 M Glycine
1 % v/v SDS
pH to 8.8

12% SDS-PAGE Gel Solution
375 mM Tris-HCl
0.1 % v/v SDS
12 % v/v 29:1 Acrylamide:Bis-Acrylamide
pH to 8.8

2.3.4 Protein purification buffers

IMAC Sample Buffer
20 mM Tris-HCl
500 mM NaCl
10 % v/v Glycerol
pH to 7.4

IMAC Wash Buffer
20 mM Tris-HCl
500 mM NaCl
20 mM to 40 mM Imidazole
10 % v/v Glycerol
pH to 7.4

IMAC Elution Buffer
20 mM Tris-HCl
500 mM NaCl
200 mM Imidazole
10 % v/v Glycerol
pH to 7.4

IMAC EDTA Solution
20 mM NaH_2PO_4/Na_2HPO_4
500 mM NaCl
50 mM EDTA
pH to 7.4

Buffers stored at 4 °C.

2.3.5 Electrophoretic mobility shift assay (EMSA) buffers

10X TBE Buffer
1 M Tris base
1 M Boric acid
20 mM EDTA
pH to 8.3

10 % EMSA Gel
10 % v/v 29:1 Acrylamide:Bis-Acrylamide
1X TBE Buffer

5X EMSA Loading Buffer
0.25x TBE Buffer
34 % Glycerol
0.2 % w/v Bromophenol Blue

EMSA Binding Buffer
20 mM HEPES
1 mM EDTA
10 mM $(NH_4)_2SO_4$
1 mM DTT
0.2 % v/v Tween 20
30 mM KCl
pH to 7.6

Binding buffer stored at 4 °C. EMSA Loading Buffer stored at -20 °C.

2.3.6 SPR – SA-chip buffers and solutions

SPR-SA Binding Buffer
20 mM HEPES, pH to 7.4
1 mM EDTA
10 mM $(NH_4)_2SO_4$
1 mM DTT
0.05 % v/v Tween 20
150 mM NaCl

SPR-SA Regeneration Solution
10 mM Tris Base
2 M NaCl

SPR-SA Wash Solution 1
50 mM NaOH
1 M NaCl

SPR-SA Wash Solution 2
50 mM NaOH
1 M NaCl
50 % v/v Isopropanol

Buffer and Solutions stored at 4 °C.

2.3.7 SPR – NTA-chip buffers and solutions

SPR-NTA Eluent Buffer
10 mM HEPES, pH to 7.4
150 mM NaCl
50 μM EDTA
0.005 % v/v Tween 20

SPR-NTA Dispenser Buffer
10 mM HEPES, pH to 7.4
150 mM NaCl
3 mM EDTA
0.005 % v/v Tween 20

SPR-NTA Nickel Solution
50 mM Tris-HCl, pH to 7.5
0.5 mM $NiCl_2$

SPR-NTA Regeneration Solution
10 mM HEPES, pH to 8.3
150 mM NaCl
350 mM EDTA
0.005 % v/v Tween 20

Buffers and Solutions stored at 4 °C.

2.4 Protocols

2.4.1 Collecting *Streptomyces* genomic DNA

Solid media plated cultures of S. *avermitilis* and *S. sclerotialus* were grown by Dr Poon and Jacqueline Gill for their PhD and Mchem work respectively. The genomic DNA of these cultures were extracted from the sporulating colonies using the FASTDNA® SPIN Kit for Soil according to its protocol as described in the manual.

2.4.2 Polymerase chain reaction (PCR)

Target genes were amplified using a three-step 35 cycle PCR utilising Phusion® High-Fidelity polymerase. Table 2-5 shows a list of primers used for the amplification of these target genes, and Table 2-8 shows the typical Mastercycler® Nexus Thermocycler PCR program for final amplifications. GC Rich Sample Mixtures and temperature conditions for each primer were optimised using a Thermocycler epGradient Thermocycler program. All forward primers were designed with a blunt 5'-CACC overhang to facilitate cloning into a pET151/D-TOPO® vector.

Table 2-8 – Optimised three-step PCR program for *Streptomyces* sp. gene amplification.

PCR Step	Temperature (°C)	Time
1. Initial Denaturation	94	1 min
2. Denaturation	94	1 min
3. Annealing	55-65	1 min
4. Extension	72	1 min
5. Go to 2. x 25-35 cycles		
6. Final Extension	72	10 min
7. Hold	4	Hold

2.4.3 Agarose gel electrophoresis

The 1 % Agarose Gel Mixture was heated in a microwave and stirred until the agarose had dissolved, and the flask cooled under a stream of cold water for approximately 30 s. 5 μL of GelRed® Nucleic Acid Gel Stain was added to the Gel Mixture and mixed thoroughly. The solution was then poured into a horizontal gel mould and allowed to set with an appropriately sized Gel Comb inserted. The gel was then transferred into a Wide Mini-Sub Cell GT Cell, and the cell subsequently flooded with 1X TAE Buffer. The 5 μL PCR mixture was mixed with 6X Loading Dye in a 5:1 μL ratio of mixture:dye, and then aliquoted into the wells. MR FastRuler™ DNA Ladder was added to a separate well as a reference. Gels were run for 40 min at 100 V, and then visualised using a UV Transilluminator.

2.4.4 Gel extraction

Extraction of amplicons from a 1 % w/v Agarose Gel was performed as described in the GeneJET™ Gel extraction kit manual. After Gel Electrophoresis the gel was visualised under a UV Transilluminator, and regions of gel containing the desired bands were excised using a Stanley knife. The extracted gel was then submerged in GeneJET™ Binding Buffer at 1:1 w/v ratio in an Eppendorf tube. The gel was dissolved in a 60 °C water bath, and the solution transferred into the GeneJET™ purification column, which was then centrifuged. After centrifugation the flow-through was discarded, and the column was centrifuged again with additional GeneJET™ Binding Buffer, and then washed (and centrifuged) using GeneJET™ Wash Buffer. In a deviation from the protocol, the DNA was eluted using 50 μL of dH_2O instead of GeneJET™ Elution Buffer. DNA concentration was then measured using NanoDrop™.

2.4.5 *E. coli* TOP10 cloning and transformation

Cloning and transformation was carried out as described in the pET151/D-TOPO® manual; the PCR product was mixed with the TOPO® vector (covalently bound to the topoisomerase enzyme) and incubated at room temperature for 5 min. The pET151 construct was mixed with Chemically Competent *E. coli* TOP10 cells and incubated on ice for 30 min, then heat-shocked at 42 °C for 30 s. The cells were transferred into SOC media and incubated at 37 °C, 200 rpm for 1 hr. The media was then spread onto LBA-AMP plates and incubated overnight at 37 °C. From the plates five colonies were selected and cultured overnight in 10-mL LB-AMP medium (LB Medium + 0.1 % w/v Ampicillin at 100 g mL^{-1}) at 37 °C, 200 rpm.

The plasmid was then extracted using the GeneJET™ Plasmid Miniprep Kit; the cells were centrifuged, the media discarded, and the cells then resuspended in GeneJET™ Resuspension solution. The cell resuspension was mixed with GeneJET™ Lysis Solution, and then mixed with GeneJET™ Neutralization Solution. Following centrifugation, the supernatant was transferred to the GeneJET™ Spin Column and centrifuged. After discarding the flow-through, the column was twice-centrifuged with GeneJET™ Wash Solution, and then eluted using 50 μL dH_2O instead of GeneJET™ Elution Buffer. The eluted plasmids were then assessed for the presence of the inserted target amplicon using the PCR program described in the pET151 manual. The plasmid sequences were then verified using GATC Biotech Sanger sequencing. Cell cultures not immediately used were stored at -80 °C in 25 % Glycerol.

2.4.6 *E. coli* BL21(DE3) transformation and expression

The Miniprep extracts were transformed into Chemically Competent *E. coli* BL21(DE3) cells as described for *E. coli* TOP10 and left growing in LB overnight. The following morning, cultures were transferred to fresh 10 mL LB-AMP media, and incubated at 37 °C, 200 rpm until OD_{600} reached 0.5-0.8. Expression was then induced with 1 mM IPTG and the cells left to incubate either overnight at 15°C to 20 °C or between 4 hrs to 6 hrs at 37 °C. After expression, a 1 mL aliquot (or a whole culture volume if proceeding straight to purification) was centrifuged and the supernatant discarded. The pellets were then resuspended in cell lysis buffer, and the cells lysed by sonication. After another round of centrifugation, both the "soluble" cell lysate and "insoluble" pellet was collected, and the soluble fraction was syringe filtered through a 0.22 μM Minisart filter. Samples not used immediately were stored at -20 °C in 10 % v/v glycerol.

2.4.7 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis

All SDS-PAGE Gels were manually cast using the Mini-PROTEAN® Tetra Cell Handcast System; 6 mL to 8 mL of 12 % SDS-PAGE Gel solution was prepared and then mixed with 100 μL ammonium persulfate and 5 μL TEMED. The solution was immediately mixed, then quickly poured between a clamped spacer and short plate, which was topped with the appropriately-sized Gel Comb. Once the gel had set, the plates were transferred to a Mini-PROTEAN® Tetra Cell, and the cell flooded with 1X SDS-PAGE Running Buffer. The gel was then equilibrated by running at 100 V for 30 min at 4 °C. To prepare the samples, the soluble fraction was mixed with 1:1 v/v of 2X SDS-PAGE Loading Buffer, and the insoluble fraction (or whole-cell pellet if cells weren't lysed) resuspended in 1X SDS-PAGE Loading Buffer with a volume equivalent to the volume aliquoted from the cell cultures pre-cell lysis.

The samples were then boiled at 99 °C in a thermocycler for 5 min. Along with 5 µL of FastRuler™ Prestained Protein Ladder, 5 µL of the insoluble fraction and 10 µL soluble fraction were pipetted into the wells of the set SDS-PAGE gel. Still at 4 °C, the gel was run at 120 V for 10 min, followed by 200 V for 30 min. The plates were then removed and separated to recover the intact gel, which was then left in Expedeon InstantBlue™ Protein Stain for 15 min to 30 min. The Stain was then drained and replaced with dH_2O, to remove excess stain from the gel.

2.4.8 Purification of protein by immobilised metal affinity chromatography (IMAC)

Prior to purification, a 1.0x30 cm Econo-Column® Chromatography Column (with additional tap and reservoir fixtures) was packed with Ni-Sepharose™ High Performance by decanting ~5 mL of the 20 % ethanol slurry into the column, allowing the resin to settle, and then draining away the ethanol solution. The resin was then washed with dH_2O, and then pre-equilibrated with the IMAC Sample Buffer. The collected cell lysate was passed through the column, and the lysate flow-through collected. The column was then washed (and wash flow-through collected) with IMAC Wash Buffer, followed by elution with IMAC Elution Buffer which was also collected. Each collected fraction was then assessed using the SDS-PAGE protocol described previously, and protein concentration measured using the NanoDrop™.

Between purifications of different proteins or following extensive use of the column, the medium was regenerated by stripping away the nickel with the IMAC EDTA Solution, washed with IMAC Sample Buffer, followed by a dH_2O wash, and finally recharged with 0.1 M $NiCl_2$ diluted in dH_2O. Excess $NiCl_2$ was washed off the column with dH_2O, before finally being equilibrated again in IMAC Sample Buffer. If not in use, the resin was stored in 20 % ethanol solution inside the column.

2.4.9 Protein concentration and buffer exchange

Eluted IMAC fractions with a low protein concentration were concentrated with an Amicon Ultra-15 (mL) Centrifugal Filter Unit MWCO 10 kDa centrifuged at 4 °C for 15 min at 4,000-7,000 relative centrifugal force (RCF), repeating until the elution flow-through was depleted. The concentrated elution was then diluted in IMAC Sample Buffer unless otherwise stated, and then filtered an additional 2-3 times. If further concentration was required, the solution was centrifuged in an Amicon Ultra-0.5 (mL) Centrifugal Filter Unit MWCO 10 kDa at 14,000 RCF for 5 min to 15 min at 4° C. For eluted protein with a high protein concentration buffer-exchange was performed *via* protein dialysis. The elution was sealed inside a semi-permeable

MWCO 10 kDa tube, and then placed in IMAC Sample Buffer at a 1:20 ratio of elution:buffer at 4 °C. The buffer was exchanged either 1-2 times after 4 hr intervals, and then left for a final exchange overnight.

For measuring the concentration purified protein samples light absorption was measured at 280 nm using NanoDrop™ (using a blank reference of the buffer-only solution), and the protein concentration calculated using the Beer-Lambert Law ($A = \varepsilon cl$; at 280 nm $\varepsilon 280 = (nTrp \times 5500) + (nTyr \times 1490) + (nCys_{ox} \times 125)$). NanoDrop™ spectra were generated over a range of 220 nm to 350 nm; with the ratio of absorbance at 260/280 nm ($A_{260/280}$) used to assess protein purity. Ratios of 1.7-2.0, 1.9 to 2.2 and 0.6 are considered pure for DNA, RNA and protein samples respectively.

2.4.10 Hairpin dsDNA probe preparation

All self-annealing oligonucleotides used for EMSA and SPR experiments are listed in Table 2-7. Oligonucleotides were ordered as 5'-biotinylated 65-bp double-stranded sequences from Sigma-Aldrich, which were then dissolved in dH_2O and transferred into PCR tubes. The DNA was boiled at 95 °C for 5 min to denature the double helix structure, and then left for 1 hr to re-hybridise as double-stranded hairpin oligonucleotides at room temperature (Figure 2-1). The choice to use hairpin dsDNA oligonucleotides was made based on the work of Voigt *et al*[200], and has been shown by Dr Zhou to retard the movement of very small sequences through polyacrylamide gels[148].

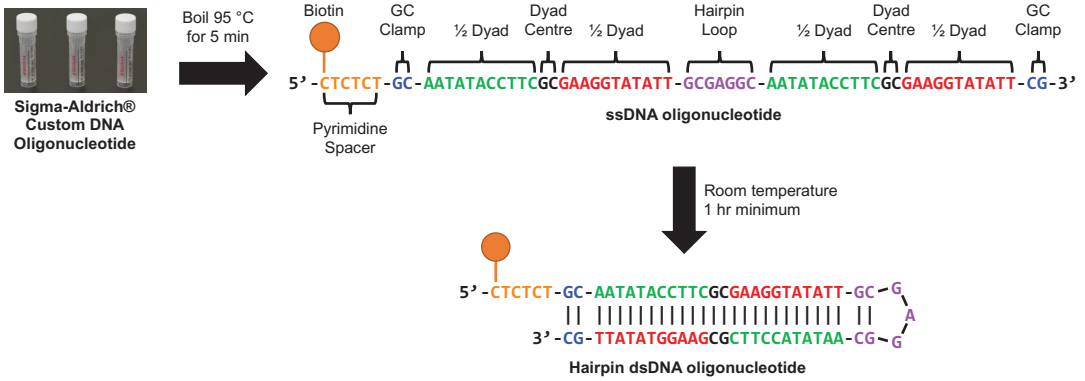


Figure 2-1 – Protocol for generation of hairpin dsDNA oligonucleotides. Each oligonucleotide features the 24-bp ARE sequence duplicated on a single strand in both its + sense (**green**) and –sense (**red**) orientation in the 5' to 3' direction for both. Each duplicated ARE is separated by a GC-rich loop (**purple**) and is capped at the ends with a GC clamp (**blue**). The 5' end of the oligonucleotide features an additional 6-bp pyrimidine spacer (**orange**) where biotin is added to the 5' end. The sequences of oligonucleotides used in gel shift assays are shown in Table 2-7.
.

2.4.11 Electrophoretic mobility shift assays (EMSA)

Gels for EMSA were cast in the same way as for SDS-PAGE, except 10 % EMSA Gel solution was used instead. After the gel had been cast, it was pre-equilibrated in 1X TBE buffer at 150 V for 1 hr at 4 °C. The protein was mixed with the oligonucleotide of interest at varying ratios in EMSA Binding Buffer, each mixture to a total volume of 20 µL, and incubated at room temperature for 15 min. For measuring derepression by AHFCA signalling molecules, the dry compound was dissolved in DMSO to 1 mM concentration, diluted in EMSA Binding Buffer to the desired concentrations, then added and left to equilibrate for a further 15 mins. 5 µL of 5X EMSA Loading Buffer was added to the samples, and the final 25 µL solutions were loaded onto the gel, and run at 80 V for 2 hrs to 4 hrs at 4 °C. To visualise ssDNA, dsDNA and RNA the gel was stained in 100 mL 1X TBE Buffer containing 5 µL of GelRed® Nucleic Acid Gel Stain intercalator for 30 min and then visualised under a UV transilluminator, and the digital image captured. Image intensity data was also collected and analysed using ImageJ.

2.4.12 Surface plasmon resonance (SPR)

Unless otherwise stated, all SPR experiments were performed at 25 °C. 5'-Biotinylated self-annealing Hairpin dsDNA oligonucleotides containing the target ARE sequences were prepared as described previously. For binding experiments with the DNA immobilised onto the chip, each experiment used an S-series SA-chip which was loaded into the Biacore T-200 apparatus. Each chip was prepped with three alternating injections of SPR-SA Wash Solution 1 and the SPR-SA Binding Buffer. After priming, solutions containing each biotinylated oligonucleotide of interest were diluted to 50 µg mL^{-1} in SPR-SA Binding Buffer and injected over flow cells (Fc) 2, 3 and 4 for 5 mins at a flow rate 10 µL min^{-1}. The reference flow cell Fc1 was flooded with the biotin-analogue biocytin to block all unbound streptavidin binding sites. Following immobilisation, the flow cells were washed with SPR-SA Wash Solution 2. Protein interactions with the immobilised DNA was then measured by injection of the protein (buffer exchanged into SPR-SA Binding Buffer) over each flow cell at a flow-rate of 30 µL min^{-1} for 150 s.

To measure dissociation, SPR-SA Sampler Buffer was subsequently injected over the chip at the same flow-rate and duration. The immobilised DNA was then regenerated (protein removed from the DNA) by injection with SPR-SA Regeneration Solution for 30 s at the same flow-rate. The chip was then buffer equilibrated, and the next cycle run. Each protein concentration sampled was replicated 3 times. To track machine consistency, samples were assayed in random order, and a buffer blank containing only SPR-SA Binding Buffer was

assayed every 5-10 cycles. All sensorgrams collected were then double blanked; first the SPR binding buffer blank average was subtracted from each data-series to negate buffer interactions, then data collected from reference Fc1 were subtracted from Fc2, Fc3 and Fc4 to negate non-specific interactions with the chip surface. For determination of thermodynamics, the protocol is outlined as above, but assayed at 7 °C, 12 °C, 18 °C, 25 °C, 32 °C and 40 °C.

For measuring inducer activity, the compounds were dissolved in SPR-SA Binding Buffer to the desired concentrations, and each inducer was then premixed with the protein at room temperature and incubated for 15min to 30 min, and binding activity then measured relative to protein binding in the absence of any inducer as a control. The desired concentrations were twice what was indicated in the experiments to account for dilution as a result of mixing.

For screening DNA oligonucleotides and inducers binding directly to the protein, an S-Series NTA-chip was loaded into the Biacore T-200 and equilibrated with 50 mM Tris-HCl pH 7.5 buffer. Following this SPR-NTA Nickel Solution was injected over the chip surface. All non-bound Nickel was washed off the chip using the SPR-NTA Dispenser Buffer. Next, the 6xHis-tagged protein which had been buffer-exchanged into the SPR-NTA Eluent Buffer was then diluted to 200 nM in more SPR-NTA Eluent Buffer and then immobilised onto the chip by injection for 60 s at 30 μL min^{-1}. To determine protein chip surface stability, SPR-NTA Eluent Buffer is injected for a further 60 s at the same flow rate. Each DNA probe, which had also been diluted in the SPR-NTA Eluent Buffer, was then injected over the chip for 120 s to assay association. SPR-NTA Eluent Buffer was then injected for 120 s to measure dissociation. After dissociation, the NTA-chip was injected with the SPR-NTA Regeneration solution to strip Ni^{2+} off the chip, and then washed with SPR-NTA Eluent Buffer, ready for another cycle.

2.4.13 Crystallographic screening

Morpheus®, PACT *premier*, JCSG-Plus™ (MD1-40), ProPlex™, Wizard Classic I and II and MIDAS™ crystal screening kits where used to screen initial crystallisation conditions using a high-throughput 96-well set-up on the Mosquito LCP, utilising the sitting drop vapour method to observe crystallisation over several weeks using drop ratios of 1:1 (100 nL crystallant and 100 nL protein; unless otherwise stated), with 50 μL of reservoir solution. Crystallant conditions can be found in the relevant documentation on the Molecular Dimensions website. Potential hits were scaled up manually on 24-well plates at room temperature, using the hanging drop vapour method technique to crystallise protein at drop ratios of 1:1 μL, 1:2 μL and 2:1 μL crystallant:protein. Conditions that successfully scaled were further optimised by manual preparation of crystallant hits.

2.4.14 Bioinformatics

The Multiple EM for Motif Elicitation (MEME)-suite was used for bioinformatic analyses in this project, a collection of tools which includes MEME, Tomtom, Find Individual Motif Occurrences (FIMO) and Motif Alignment and Search Tool (MAST). MEME is a motif discovery tool for identifying ungapped motifs (recurring, fixed-length) using statistical modelling to identify patterns amongst sets of DNA, RNA or protein sequences. Tomtom is used to align and measure statistical similarity between motifs. FIMO takes either an output of MEME or a manual input of a motif and parses it through a submitted sequence database to identify statistically-significant (as defined by the user) occurrences of the submitted motif. MAST can be used to compare multiple motifs with a set of sequences and scores each sequence against the submitted motifs.

To generate and identify sequence motifs in ArpA-subfamily repressor recognition sites, confirmed binding sites from literature were parsed through MEME to generate a motif. FIMO was used to identify novel occurrences of the motif in sequences as indicated in the relevant chapters, using the MEME-generated motif. Tomtom was used to compare sequences bound by ArpA-subfamily repressors. MAST was used for enrichment of select motifs in sequences predicted from FIMO, using motifs generated in MEME.

2.4.15 *In silico* docking

The 1.5 Å resolution structures of native MmfR and MmfR:AHFCA2 previously solved by Dr Rea (personal correspondence) were used as the template for homology modelling of MmfR and MmyR orthologue structures using the Phyre[2] web engine. A Ligand .pdb library of AHFCAs, AHFCA analogues and other small molecules was generated using ChemAxon MarvinSketch (available in the Electronic Supplementary Materials). Docking of all ligands were performed in the AutoDock Tools suite, utilising AutoDock Vina (ADV) as the primary algorithm method specifically for the docking of flexible ligands to rigid protein structures. Each docking was replicated five times, with the final docking scores (in kcal mol^{-1}) being the mean scores for correct conformation selected by analysis in PyMOL.

For docking of MmfR to DNA a flexible DNA structure was generated using 3D-DART, and MmfR docked to it using the HADDOCK web service. For comparison with CprB binding to DNA (PDB:4PXI) the DNA structures contained in each pdb file were aligned using the "cealign" command in PyMOL and validated by checking the alignment of the dyad centre nucleotides of each sequence. The docked structure can be found in the Electronic Supplementary Materials.

2.5 Experimental principles

2.5.1 pET151/D-TOPO® cloning system

The pET151/D-TOPO® cloning system (Figure 2-2) utilises a single-step cloning reaction catalysed by a topoisomerase enzyme, rather than by typical restriction enzyme-based methods. Genes are inserted into a TOPO site under the control of a T7 promoter, which enables transcription only when transformed into organisms encoding the T7 RNAP, such as *E. coli* BL21(DE3). The forward primers are modified with a 5'-CACC blunt-end extension, which enables recognition of the amplicon by D-TOPO® enzyme and subsequent oriented insertion into the TOPO insertion site. Downstream of the T7 promoter is a *lacO* operator site for the repressor lacI. Induction of cell cultures with the lactose homologue IPTG enables transcription of the open reading frame (ORF). The genes would be acquired by amplification with PCR using the *Streptomyces* genomes as a template. The gene is expressed with an extended N-terminal tag which includes a 6xpolyhistidine tag (His_6) utilised in IMAC, a V5 epitope for western blot analysis, and a TEV site for cleavage by TEV Protease. The expressed protein would then be purified using IMAC.

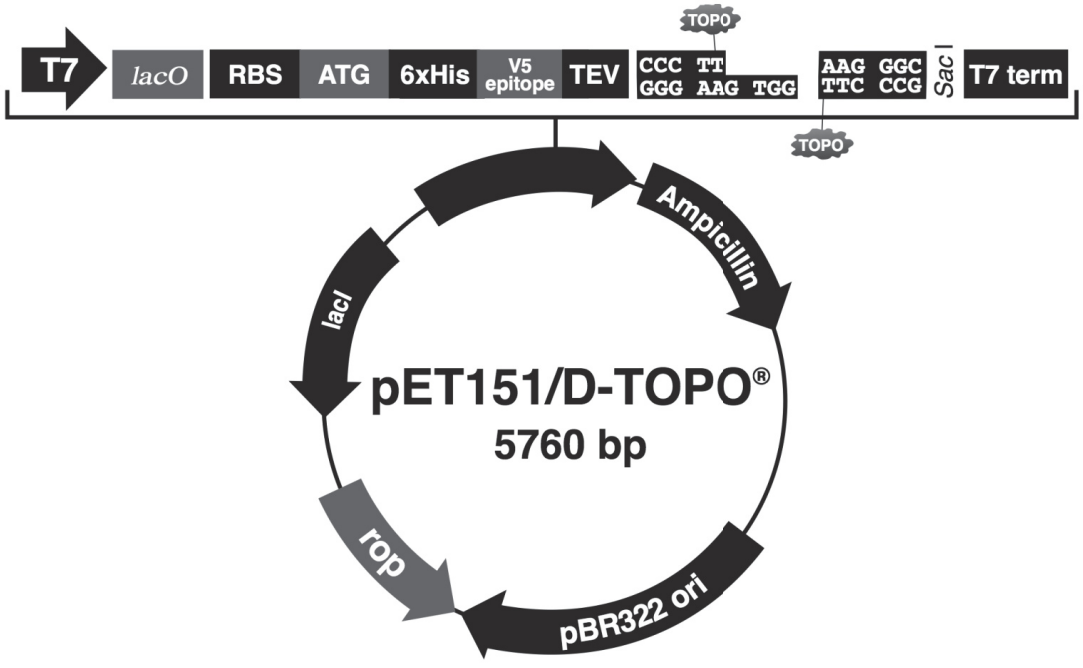


Figure 2-2 – Plasmid map of the pET151/D-TOPO®-cloning vector taken from the pET151 manual. *lacO* is the LacI operator (allowing induction by IPTG), RBS is the Ribosome binding site, ATG is the start codon, and TEV is the TEV protease cleavage site for the expressed recombinant protein. Directional cloning is performed by incubation of the amplified gene (amplicon) with topoisomerase, which requires incorporation of a 5'-CACC overhang into the 5' end of the amplicon sequence. Transcription of the cloned gene is under the control of a T7 promoter, which enables transcription only in an expression host that produces T7 RNAP, such as *E. coli* BL21(DE3).

2.5.2 EMSA principles

Gel shift assays are a DNA or RNA-based experimental assay typically used to characterise the formation of protein-DNA or protein-RNA complexes (Figure 2-3). DNA is loaded into the well of the polyacrylamide gel; when an electric current is run across the tank the negatively-charged DNA/RNA is repelled through the gel from the negatively-charged cathode towards the positively-charged anode. Shape, charge and mass all factor into that affect how effectively molecules move through the gel. When a protein or other macromolecule interacts with the DNA/RNA, it affects all three of these properties; the physical size and mass of the complex increases, and overall charge polarity is generally reduced. Thus, the movement of DNA/RNA through the gel is retarded, causing bands with different migratory speeds to appear in dose-dependent fashion on the gel. This is referred to as a 'gel shift'. If enough protein is added, the band will be completely shifted so that the original band of free DNA/RNA no longer appears. It is not uncommon to find multiple complexes formed with proteins that bind at a stoichiometry greater than one minimally binding unit (in the case of TFRs, a dimer). Then, by adding an effector which regulates the activity of the protein, the protein can be released from DNA/RNA in a similarly dose-dependent fashion (in the case of typical TFRs). These observed shifts can be quantified in gel analysis software, given that the captured image of the gel is of sufficient quality.

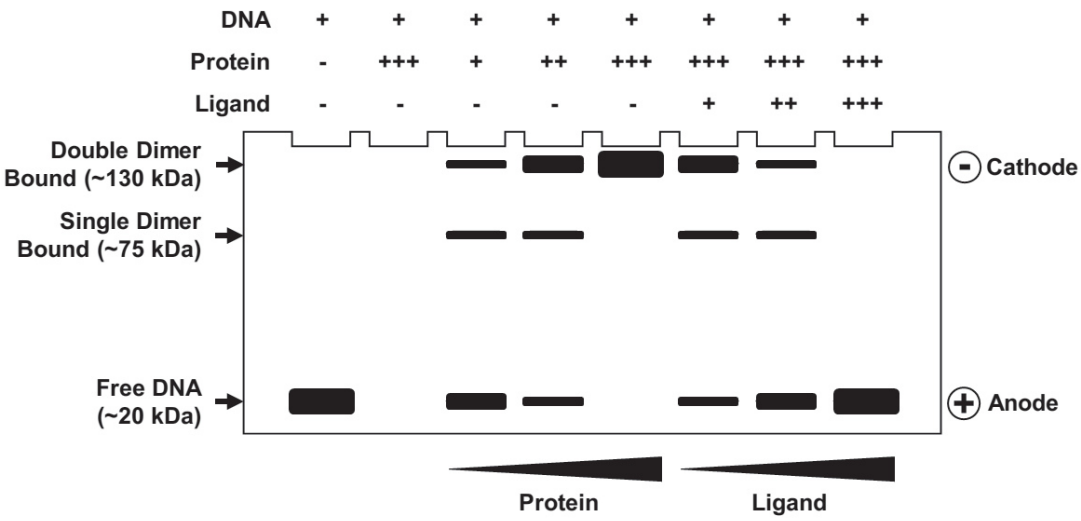


Figure 2-3 – Basic gel shift assay protocol for investigating DNA-binding and ligand-binding activity of transcriptional repressors. Molecular weights are assumed for a self-annealed dsDNA hairpin oligonucleotide for DNA, and MmfR for protein:DNA complexes.

2.5.3 SPR theory

2.5.3.1 Terminology and general principles

Table 2-9 – Important SPR-related terminologies.

Term	Definition
Ligand	Molecule to be immobilised onto the chip surface. Not to be confused with ligand defined in other contexts.
Analyte	Molecule injected over the ligand-immobilised chip to measure binding between ligand and analyte.
RU	Response units; 1,000 RU is defined as a difference of 0.01° between θ_1 and θ_2 (approximate to 5 ng mm^{-2} of binding for most proteins).
R	Response in RU.
R_L	Response in RU of injected ligand-binding to the chip.
R_A	Response in RU of injected analyte binding to ligand.
R_{eq}	Fitted R_A in RU for a given concentration when the reaction is in steady state equilibrium.
R_{max}	Maximum R_A in RU at saturation of ligand calculated during model fitting.
TR_{max}	Theoretical maximum R_A in RU at saturation of ligand at saturation calculated based on 100 % functionality of ligand chip.
Binding Level (*BL*)	Value of R_A typically interpreted as the maximum observed binding for a given cycle. It is taken 4 s prior to start of the dissociation phase by default.
RI	Bulk refractive index contribution used in model fitting.
C	Concentration of analyte in solution.

SPR is a physical phenomenon of electromagnetic waves that can occur when plane-polarised light fired through a glass prism coated with a metal film (typically gold) at the correct angle (incidence angle > critical angle) undergoes total internal reflection as it hits the prism-gold interface. The reflected light is captured by a detector that measures the angle and intensity of the reflected light. Free electrons in the gold film undergo periodic oscillation, a waveform referred to as a plasmon. Plasmons on the surface of the gold film generate an electrical field ~300 nm into both the buffer and the gold layer, referred to as an evanescent wave. Single-wavelength multi-angle photons hitting the surface at the correct resonance angle (θ_1) with a momentum equal to the momentum of the oscillating plasmons results in optical coupling. The absorbed photons are no longer reflected onto the detector, resulting in a measurable decrease in light intensity. Alterations in composition on the gold surface within the range of the evanescent wave result in a change in the momentum of the surface plasmons proportional to the change in mass on the surface, altering the refractive index of the interface. As a result,

SPR no longer occurs at θ_1, but now occurs at a different resonance angle, θ_2. The change in angle between θ_1 and θ_2, is expressed as a measurement of binding in response units (RU).

In Biacore systems, chips housing the gold-coated prism have a 100-nm carboxymethylated (CM5) dextran layer incorporated onto the gold surface, which enables immobilisation of molecules onto the chip *via* coupling with the dextran surface. Each chip typically houses four flow cells, the first of which is typically used for reference subtraction. For testing of a biological interaction, the first interactant, defined as the ligand, is injected over the chip using the appropriate method for the chip type. For the experiments in this thesis, both streptavidin (SA) chips, and nitrilotriacetic acid (NTA) chips were utilised. Prior to purchase, streptavidin was first pre-immobilised onto the chip *via* amine coupling with the dextran surface, allowing simple, highly-stable capturing of any biotinylated molecules. NTA-chips uses the same principle as IMAC; first $NiCl_2$ was injected onto the chip, then His_6-tagged proteins are injected to chelate with the captured nickel ions. Once immobilised, the second interactant, referred to as the analyte, was injected over the chip surface, and the increase in mass as a result of the interaction between analyte and ligand is measured in RU between the θ of the immobilised ligand and the θ of the ligand:analyte complex. From this data, the affinity, kinetics, stoichiometry, thermodynamics and other useful variables could be discerned for the interaction of the two molecules.

2.5.3.2 Model fitting

When fitting either affinity or kinetic models, models can be fitted either locally or globally. Local fittings are used to fit each individual curve within a dataset, while global fitting utilises all curves for fitting. Often models use a mixture of both; fitted variables are defined as either local parameters or global parameters. To determine a good model fit, visual inspection of the fit, Chi^2, R_{max} and model residuals can be used to assess the quality of the model fitting to the data. Chi^2 (RU^2) describes how well the model fits the observed data:

$$Chi^2 = \frac{\sum_1^n (R_f - R_x)^2}{n - p}$$

Where R_f is experimental R, R_x is the fitted R, n is the number of data points, and p is the number of fitted parameters. Chi^2 is the mean squared residual for the difference in RU between the experimental data and the fitted curve. Chi^2 can also be measured as a percentage of R_{max}: by rule of thumb Chi^2 should be < 10 % the value of R_{max} for a strong fit, but one must be careful with interpretation as very poor fits can produce very high estimates of R_{max},

resulting in much lower percentage values of Chi^2. R_{max} should be checked to ensure that it is not significantly lower or higher than assayed concentrations near saturation of the ligand, or if not at this point then compared to the theoretical maximum response (TR_{max}). Features of a poor model fit include poor visual fitting, a high Chi^2/percentage of R_{max}, and large residuals compared to the response of the overall data. For a good fit, if the total mean of residuals for each curve is below $\sqrt{Chi^2}$, then Chi^2 becomes the mean square of the signal noise. A strong model fit will also have a random residual pattern, whereas a poor fit will show systematic deviation from the data; this often implies that the applied model does not adequately explain the binding response. If the biological mechanism is known, then other factors affecting a poor fit include poor analyte or ligand preparation: misfolded protein, mis-prepared samples or impurities in the buffer can often contribute to secondary non-specific binding. This can be exacerbated as the resolution of the SPR equipment increases.

2.5.3.3 Steady state affinity modelling

For an assumption of steady state equilibrium following a 1:1 Langmuir model (Langmuir isotherm), the net rate of complex formation is zero:

$$\frac{dR_{eq}}{dt} = k_a C R_{max} - (k_a C + k_d) R_{eq} = 0$$

$$\therefore k_a C R_{max} = (k_a C + k_d) R_{eq}$$

Rearranging and substituting $\frac{k_d}{k_a} = K_D$, the equation fitted is $R_{eq} = \frac{C \times R_{max}}{K_D + C} + RI$.

Where the initial fitting value for R_{max} is defined as Y_{max}, RI (bulk refractive index contribution) $= \frac{Y_{max}}{5}$, and $K_D = \frac{X_{max}}{10}$. R_{max}, C (analyte concentration) and RI are local parameters, while K_D is a global parameter. For measuring affinity and kinetics, the amount of ligand immobilised on the chip should be low enough such that adequate saturation of the ligand can be achieved; the analyte concentration should cover a range more than twice the value of the determined K_D.

2.5.3.4 Kinetic modelling

For kinetics, an additional term for mass transfer, the mass transfer constant k_t (RU M^{-1} m s^{-1}) is included in model fitting. Mass transfer describes the movement of analyte in buffer towards the chip surface before interacting with the immobilised ligand. The flow rate-independent component of the mass transfer constant, referred to as *tc* (RU M^{-1} $s^{-2/3}$ $m^{-1/3}$) is fitted globally to the data, then flow rate is factored in for k_t determination:

$$tc = \frac{\left(0.98\left(\frac{D}{h}\right)^{2/3} \times \left(\frac{f}{0.3 \times w \times l}\right)^{1/3}\right) \times mW_{analyte} \times 10^9}{f^{1/3}} = \frac{k_m \times mW_{analyte} \times 10^9}{f^{1/3}} = \frac{k_t}{f^{1/3}}$$

$$\therefore k_t = tc \times f^{1/3}$$

Where *D* is the analyte diffusion coefficient, *f* is the volume flow rate, *h*/*w*/*l* are the flow cell dimensions, and k_m is the mass transfer coefficient (m s^{-1}). For 1:1 Langmuir kinetics ($A + L \rightleftharpoons AL$), the fitted rate equations are as follows:

$$\frac{d[A_{surface}]}{dt} = k_t \times \left([A_{solution}] - [A_{surface}]\right) - \left(k_a[A_{surface}][L] - k_d[AL]\right)$$

$$\frac{d[L]}{dt} = -\left(k_a[A_{surface}][L] - k_d[AL]\right)$$

$$\frac{d[AL]}{dt} = k_a[A_{surface}][L] - k_d[AL]$$

$$R_A = [AL] + RI$$

With the following assumptions:

$[\boldsymbol{A_{solution}}] = \boldsymbol{C}$; concentration of analyte is given as concentration of analyte in solution.
$\boldsymbol{A_{surface}[0] = 0}$; when analyte concentration at the chip surface = 0, then $R_A = 0$.
$\boldsymbol{L[0] = R_{max}}$; when free ligand concentration at the chip surface = 0, then $R_A = R_{max}$.
$\boldsymbol{AL[0] = 0}$; When no complex formed, $R = 0$

RI is fitted locally, k_a, k_d, R_{max} and *tc* are fitted globally, and both *C* and *f* are constant. For accurate kinetic determination the amount of ligand immobilised onto the chip should be as low as possible to reduce effects from mass transfer and steric hindrance. The ideal R_{max} is usually taken to be 100 RU, for which the amount of ligand immobilised is optimised for assayable saturation of the ligand.

2.5.3.5 Kinetic models

Note that before fitting a model one must be careful not 'model shop' for binding mechanisms: just because a mathematical model may fit the binding data, this does not mean that the model truly reflects the biological events that are occurring. A fitting should always have either a biological basis for rationalisation (either from complimentary experimental methods or available literature on homologues), or an experimental one based on knowledge of the samples prepared. Unless no rational is made, a 1:1 Langmuir model should always be fitted to the data. If no rationalisation is made and 1:1 Langmuir does not fit, all that can be said is that the 1:1 Langmuir model does not describe the binding event occurring.

Table 2-10 below shows the four default models in addition to the 1:1 Langmuir model that are included in the Biacore T200 evaluation software. The bivalent analyte model is used for analytes which interact with the two different ligand molecules *via* two identical binding sites, such as an in antibody interactions. K_D is typically not calculated for this model, as the two sets of calculated rate constants are difficult to interpret, where it is inferred that binding to the second ligand is in some way is dependent on the formation of the complex with the first ligand. The two state reaction model, also known as the conformational change model, describes a binding reaction with a second reaction step, resulting in two distinct association and two dissociation phases. The first set of rate constants characterise the binding of the analyte itself (which produces the measured response), which usually has a rapid initial association and dissociation phase. The second set of rate constants describe a conformational state change in the complex once bound, often characterised by a second slow, association/dissociation phase.

The heterogeneous ligand model, also known as the heterogeneous surface or surface heterogeneity model, is typically used when there is a known impurity in the ligands immobilised on the chip. Similarly, the Heterogeneous analyte, also known as heterogeneous solution or solution heterogeneity model, can only be used when the known quantity of two analytes is known. These models should never be used unless a rationale is made for their use.

Several issues arise when considering even more complex kinetic models than what the developers integrated into their evaluation software. One concerns timescales and complexity of the models built. Accurate determination of binding mechanism from data fitting is dependent on measurable rates of association and dissociation, which is why steady state affinity is better suited for K_D determination for an interaction which has high association/dissociation rates. Similarly, as extended models may have many steps, accurately

distinguishing kinetics between transitional states becomes too difficult to interpret. It is anticipated that successive reaction steps follow one after another, but the reality is that the mechanisms may be more dynamic than can currently be modelled for, and approximation of mechanisms are the best that can be collected at this time.

Table 2-10 – Default kinetic models provided with the Biacore T-200 evaluation software. A denotes the injected analyte (protein in the SA-chip protocol, DNA in the NTA-chip protocol), and L denotes the immobilised ligand (surface macromolecule; DNA in the SA-chip protocol, and protein in the NTA-chip protocol). Bivalent Analyte K_D is not included in the default model.

Kinetic Model	Reaction steps	K_D Determination
1:1 Langmuir	$A + L \rightleftharpoons AL$	$K_D = \frac{k_d}{k_a}$
Bivalent Analyte	(1) $A + L \rightleftharpoons AL$ (2) $AL + L \rightleftharpoons AL_2$	
Two State Reaction	(1) $A + L \rightleftharpoons AL$ (2) $AL \rightleftharpoons (AL)$	$K_D = \frac{k_{d1}}{k_{a1}} \times \frac{k_{d2}}{k_{d2} + k_{a2}}$
Heterogeneous Analyte	$A^1 + L \rightleftharpoons A^1L$ $A^2 + L \rightleftharpoons A^2L$	$K_{D1} = \frac{k_{d1}}{k_{a1}}$ $K_{D2} = \frac{k_{d2}}{k_{a2}}$
Heterogeneous Ligand	$A + L^1 \rightleftharpoons AL^1$ $A + L^2 \rightleftharpoons AL^2$	$K_{D1} = \frac{k_{d1}}{k_{a1}}$ $K_{D2} = \frac{k_{d2}}{k_{a2}}$

Another problem is that as a rule the more complex the model becomes, the easier it becomes to overfit the model to the data, due to the increased number of variables than can be manipulated. This why these additional models are subject to the same considerations as posed earlier regarding model shopping and overfitting to binding data. SPR does not necessarily identify what the correct model is, but it may be used to eliminate obvious discrepancies between models as a way of negating specific models. When the differences between models become minute (multiple models having near equivalent viability), alternative means of determining binding mechanisms should be investigated.

2.5.3.6 Stoichiometry

None of the kinetic models listed above can indicate stoichiometry of binding, except in the case of the bivalent analyte model (with biological inferences). However, both steady state and kinetics modelling can be used to estimate valency based on the determined value of R_{max}. For the binding of an injected analyte to a chip-immobilised ligand, TR_{max} (assuming 100 % functionality of immobilised ligand), can be calculated thus:

$$TR_{max} = \frac{(R_L \cdot mW_A \cdot Valency_L)}{mW_L}$$

Where R_L is ligand response after immobilisation, and R_A is response of analyte binding to ligand (assuming no interactions with the chip). Because $R \propto$ Mass, $\frac{RU}{mW}$ becomes equivalent to $\frac{mass}{mW}$, which gives the mol-equivalent unit of RU Da^{-1}. Knowing how much RU is attributed to an analyte or ligands molecular weight (mW) allows us to determine their stoichiometric ratio (SR) from the ratio of their response. The ratio of binding of analyte to immobilised ligands can be determined using the rearranged equation:

$$Valency_L = \left(\frac{R_A}{mW_A}\right) \div \left(\frac{R_L}{mW_L}\right)$$

Where R_A may be substituted with R_{max} to determine $Valency_L$ at R_{max}.

2.5.3.7 Thermodynamics

With the acquisition of the Biacore T-200, thermodynamics studies of the interactions between macromolecules can be analysed at a temperature range of 4 °C to 45 °C, down to 20 °C below ambient temperature. The main purpose of this assay was to determine the proteins stability over a range of temperatures, but it may also be used to characterise reaction steps for fitted kinetic models. Even without running a specific thermodynamic experiement, standard Gibbs free energy ($\Delta G°$) and the equilibrium constant are related by the Nerst equation, so $\Delta G°$ can be determined from the equation $\Delta G^o = -RT \ln K_A = RT \ln K_D$, where T is the absolute temperature in Kelvin and R is the ideal gas constant (8.314 J mol^{-1}).

A Van't Hoff graph is a plot of either $\ln K_A$ or $\ln K_D$ against $1/T$. The graph is designed for estimations of standard change in enthalpy, ΔH°, and standard change in entropy, ΔS°, by taking the logarithm of the Gibbs free energy equation $\Delta G° = \Delta H° - T\Delta S°$, resulting in the equation $\ln K_A = -\frac{\Delta H°}{RT} + \frac{\Delta S°}{R}$. Assuming ΔH° and ΔS° are constant with temperature, for a linear fit of the form $y = mx + c$, the slope $m = -\frac{\Delta H°}{R}$ and the y-intercept $c = \frac{\Delta S°}{R}$, and therefore $y = -\frac{\Delta H°}{R}x + \frac{\Delta S°}{R}$. For a graph plotted with $\ln K_A$ if $m > 0$ then $\Delta H^\circ < 0$, and the binding reaction is therefore exothermic. When $m < 0$, $\Delta H^\circ > 0$, and therefore the reaction is endothermic.

The Eyring equation is a transition state analysis method relating reaction rate and temperature to the activation energy of any reaction. The equation is the form $k = \frac{k_B T}{h} \cdot e^{-\frac{\Delta H^{\ddagger}}{RT}} \cdot e^{\frac{\Delta S^{\ddagger}}{R}}$, where k is the rate constant, k_B is the Boltzmann's constant, h is the Planck's constant, $\Delta H^{\ddagger}$ is the enthalpy of activation, and $\Delta S^{\ddagger}$ is the entropy of activation. By plotting $\ln(k/T)$ against $1/T$, the resulting graph has a slope of $-\frac{\Delta H^{\ddagger}}{R}$, and a y-intercept of $\frac{\Delta S^{\ddagger}}{R} + \ln\frac{k_B}{h}$, which can be used to derive $\Delta H^{\ddagger}$, $\Delta S^{\ddagger}$, $T\Delta S^{\ddagger}$, $\Delta G^{\ddagger}$, and activation energy, E_a ($E_a = \Delta H^{\ddagger} + RT$).

2.5.3.8 Measuring effector activity by inhibition of TFR DNA-binding

To measure the ability of effectors to bind to and affect the proteins DNA-binding activity, an assay was designed based on EMSA protocols to measure the relative inhibitory activity (IC_{50}) of the signalling molecules. The same methodology has since been used to assay the inhibitory ability of 3Och-CoA and 4-BNC-CoA on the TFR *M. tuberculosis* KstR[(201)]. The effectors were premixed at different concentrations with the same concentration of protein for 15 min. The mixed solutions were then injected over DNA immobilised onto an SA-chip, and the binding level (*BL*; taken as response of analyte 4 s before dissociation) at each concentration of effector was determined. The *BL* value for each concentration of effector was then measured as the *BL* value relative to the uninhibited protein *BL* value, from which the IC_{50} was derived.

Binding of small molecules was also assessed through Binding Potential (B_{max}/K_D; *BP*), a saturation analysis technique used in pharmacokinetics to describe the affinity of a drug for its target receptor, based on total density of receptors in a system. To determine *BP* the data collected for IC_{50} determination was y-axis inverted, and the data fitted with a saturation curve, from which B_{max} and K_D are observed. In an accurate fitting of both *BP* (M) and IC_{50} (M^{-1}), the values of one is the reciprocal of the other. Determination of *BP* was performed to compare activity of protein:effector interactions *in vitro* as has been performed *in vivo* (in *Streptomyces*) by Dr Styles, who measured *BP* to quantify the activity of AHFCAs against MmfR[(149)].

Although this was primarily an inhibition experiment, a combination of inhibition and derepression was expected to occur, as any effector not bound to protein when premixed may also bind with free protein that saturates the DNA over the course of the experiment, assuming there was enough time at the programmed flow rate (30 μL min^{-1}) to interact.

2.5.3.9 Direct binding of effectors to TFRs

By immobilising the protein onto an NTA-chip, affinity and kinetics of protein:effector interactions can be measured directly, independent of the effector's biological activity. It was assumed that the effectors ability to inhibit the DNA-binding activity of TFRs is directly proportional to the affinity of binding, such that K_D should be approximate to IC_{50}. Because of limitations in signal resolution, earlier models of Biacore instruments were only reliable to assay molecules with a minimum weight of ~200 Da. However, the Biacore T-200 provides enough resolution over signal noise to determine the binding of each AHFCA. In addition, assuming AvaL1 immobilises on the chip as a homodimer, this provides two sites for each ligand to bind, therefore the actual bound molecular weight per immobilised protein is doubled. The *mW* of assayed ligands ranged from 182 Da (MmA) to and 244 Da (SCB1).

2.5.4 Statistical analysis

Statistical *t* tests were applied to much of the experimental data collected for this thesis, with significance expressed as a calculated *p* value. The *p* value indicates the probability of the observed results occurring if the null hypothesis were true. For most of the statistical tests performed for comparing two or more sample populations the null hypothesis states that all sample populations have a mean difference of 0. For all *t* tests an alpha level of 0.05 was used as a determinant of significance, meaning that if $p > 0.05$ then the evidence against the null hypothesis was not sufficient to reject it, and if $p < 0.05$ then the evidence was strong enough to reject the null hypothesis. The directionally of the differences in mean values was not tested for, so all *t* tests performed are two-tailed (unspecific for directionality of change). Each *p* value is provided where possible with a *t* test statistic and the degrees of freedom (*df*). The *t* value is the difference relative to the variation in the sample data (alternatively defined as the signal-to-noise ratio), and *df* (indicated in parentheses) the number of parameters that may vary independently.

For samples with two sample populations three different *t* tests are used;). A paired Student's *t* test is applied when two observations under different conditions are made for same subject ($df = n$ pairs – 1). When sample populations were of difference sample sizes a Levene's test was first performed to first determine scedasticity (equality of variance between sample populations), given with a *W* test statistic, and a null hypothesis of homoscedasticity (variance is equal between populations An unpaired Student's *t* test was performed when the population variances were homoscedastic, and an unpaired Welch's *t* test was performed when the population variances were heteroscedastic; in both tests $df = n$ samples from all populations –

2. For comparison of more than two sample populations ANOVA (analysis of variance) were performed (with a given *F* test statistic), with *post hoc t* tests performed as required. A Levene's test was also used for ANOVA with unpaired samples. Additional *t* tests are performed to quantify the significance of correlations between two variables, with the null hypothesis stating that there is no correlation. For linear relationships a parametric Pearson's correlation coefficient is used, and for either non-linear relationships or correlating discrete variables (*i.e.* ranks) a nonparametric Spearman's correlation coefficient (r_s) was used. For this *t* test $df = n$ pairs – 2, and *t* values are not reported.

When grouping together error values between different sample populations, Satterthwaite approximation is used to determine the pooled error, which accounts for difference variances in each sample populations. All statistical analyses are performed using the data analysis tools available in Microsoft Excel.

2.5.5 Cooperativity analysis: Hill coefficient

The Hill coefficient, n_H, derived from experiments describing oxygen binding to haemoglobin[202], is used as a measure of cooperativity between ligands binding to a receptor with multiple ligand-binding sites, defined as the fraction of a receptor with more than one ligand-binding site saturated by the ligand as a function of ligand concentration. n_H is equal to the slope of a linear regression through a plot of log [L] against log $\left(\frac{\theta}{1-\theta}\right)$, where θ is the fraction of ligand bound to the receptor and 1 is the maximum observable binding. When $n_H = 1$ ligand-binding to each site is independent, whereas $1 < n_H < n$ indicates that each ligand bound promotes further ligand-binding to the receptor up to the maximum number of ligand-binding sites, *n*. It is also possible that $0 < n_H < 1$, indicating that each ligand bound prevents further ligand-binding at other receptor sites. For TFRs and the experiments described in this thesis this can represent two interactions; one describing the protein:DNA interaction where DNA functions as the receptor and the TFR as the ligand, and through effector:protein interactions with the TFR functioning as the receptor.

3 *In vitro* functional & biophysical studies of recombinant MmfR/MmyR orthologues

3.1 Aims and Strategy

The role of AHFCA and pseudo-AHFCA receptors in coordinating AHFCA-regulated biosynthetic systems both in tandem with each other receptor and with other biological pathways is not yet fully understood. Investigating their function *in vivo*, while yielding crucial biological information, has limited use in characterising the underlying mechanisms of the receptors. To that end, the desire of this thesis is to characterise the activity of both DNA-binding and ligand-binding activity of these receptors *in vitro*. The aim of the work described in this chapter is to characterise the basic *in vitro* expressibility and activity of the AHFCA/pseudo-AHFCA receptors in *S. avermitilis* (AvaL1/AvaL2) and *S. sclerotialus* (SclM1/SclM4), for which more accurate functional information can be derived from follow-up work in subsequent chapters.

Initial protocols for expression and purification were designed based on prior *in vitro* expression of *S. coelicolor* A3(2) MmfR[139] and *S. venezuelae* SgnR[148]; *S. coelicolor* A3(2) MmyR is the only pseudo-AHFCA receptor to have been previously expressed, and has not yet been solubilised *in vitro*[139]. Both MmfR and SgnR expressed using a pET151 construct are demonstrably soluble, but only when expressed at incubation temperatures ≤ 15 °C after induction with IPTG. The recombinant proteins when purified were also active for both DNA-binding and ligand-binding assays performed using gel shift assays, so this method would also be used as an activity test of the recombinant proteins successfully solubilised and purified.

From the outset genomic isolates of *Streptomyces spp.* cultured in the C126a Microbiology lab the target genes would be used to amplify our genes of interest, which would then be cloned into pET151 plasmid vectors for propagation in *E. coli* TOP10 cells *via* incorporation of a topoisomerase 5’ orientation site (5’-CACC-3’). The plasmid would subsequently be transformed into *E. coli* BL21 DE3 cells. The recombinant protein encoded by the pET151 vector subsequently produced by *E. coli* BL21 DE3 cells would then be purified from cell culture (if soluble), and activity assessed by gel shift assays. Target AREs thought to be the native targets of these repressors would be bioinformatical predicted using Find Individual Motif Occurrences (FIMO), a sequence motif scanning tool available as part of The Multiple Em for Elucidation of Motifs (MEME) Suite. The initial 24-bp search motif was built in MEME from confirmed sites identified in the *S. coelicolor* A3(2) and *S. venezuelae* AHFCA-regulated gene clusters confirmed to be bound *in vitro* by MmfR and SgnR.

To further elucidate the biochemical mechanism of the purified repressors, crystallisation trials for each repressor were performed; with each repressor both with and without their ligands present. Given the high sequence identity of these repressors to MmfR (42.5 %, 53.5 % and 52.2 % sequence identity for AvaL1, SgnR and SclM1 respectively), for which a crystal structure was already available, it was anticipated that the phase problem for diffraction of any crystals produced could be solved *via* molecular replacement, rather than by multi-wavelength anomalous dispersion (MAD). Therefore, selenomethionine-substituted proteins were not produced.

3.2 Preparation of *S. avermitilis* AvaL1/AvaL2 and *S. sclerotialus* SclM1/SclM4

3.2.1 PCR Amplification

Successful extraction of genomic DNA from *S. avermitilis* and *S. sclerotialus* was confirmed by PCR using an actinomycetes-specific $F243_F$[203] forward primer and a universal $EUB518_R$ reverse primer to amplify 16S rDNA using a three step PCR program with an annealing temperature of 55 °C (not shown). Figure 3-1 shows agarose gels for PCRs performed to amplify the MmfR/MmyR orthologue genes from both *S. avermitilis* (*avaL1*/*avaL2*) and *S. sclerotialus* (*sclM1*/*sclM4*). For all target sequences an initial standard set of primers (F and R for forward and reverse primers respectively) were designed with melting temperatures (T_m) of approximately 50 °C to 60 °C, annealing lengths of 18-22 bp, and minimal percent GC content, with later alternate primers being designed during PCR optimisation (Table 2-5, Chapter 2.2). Amplification of the *S. avermitilis* orthologue genes were unsuccessful when using the genomic DNA extract due to assumed non-specific amplification of GC rich sequences, so a pESAC-13A-19K construct encoding the entire *S. avermitilis* azoxy compound biosynthetic gene cluster was instead used as a template. From this template *avaL1* could be amplified using all combinations of primers, with $avaL1_F$ and $avaL1_R$ being used for the final amplifications of *avaL1*.

Amplification of *avaL2* using the primers $avaL2_F$ and $avaL2_R$ resulted in largely non-specific amplification of DNA, attributed to the GC-rich content of the 3' end of the gene (the $avaL2_R$ primer having 89 % GC content). To reduce non-specific amplification two alternative reverse primers were designed; the shortened $avaL2_{RS}$ reverse primer (15 annealing base-pairs, 87 % GC content, and the $avaL2_{RD7}$ reverse primer (15 annealing base pairs, 72 % GC), which encodes an alternative stop codon resulting in truncation of the final seven codons. PCR with $avaL2_{RS}$ resulted in both specific amplification of the *avaL2* gene and non-specific

amplification of DNA, whereas *avaL2*$_{RD7}$ significantly reduced non-specific amplification. PCR using a shortened forward primer *avaL2*$_{FS}$ showed little difference in amplification compared with the *avaL*$_{2F}$ primer. *avaL2*$_{FS}$ and *avaL2*$_{RS}$ were used for the final amplifications, with the specifically amplified *avaL2* band carefully excised from the gel.

Amplification of *sclM1* failed to yield any band that correlates with the size of the expected DNA product. Possible bands can be observed for *sclM4* amplification with the *sclM4*$_{F}$/*sclM4*$_{RD1}$, *sclM4*$_{FL60C}$/*sclM4*$_{RD}$ and *sclM4*$_{FL60C}$/*sclM4*$_{RS}$ primer combinations in the 400-bp to 850-bp region, but not at an intensity that would indicate it was the correct target. Rather than attempting further optimisation of PCR, GeneArt™ Strings™ DNA Fragments encoding *E. coli* codon-optimised ORFs of the *S. sclerotialus* orthologues were purchased (Figure 8-1, Appendix B), designated *sclM1*$_{CO}$ and *sclM4*$_{CO}$. These sequences were not amplified by PCR but directly cloned into the pET151 vector.

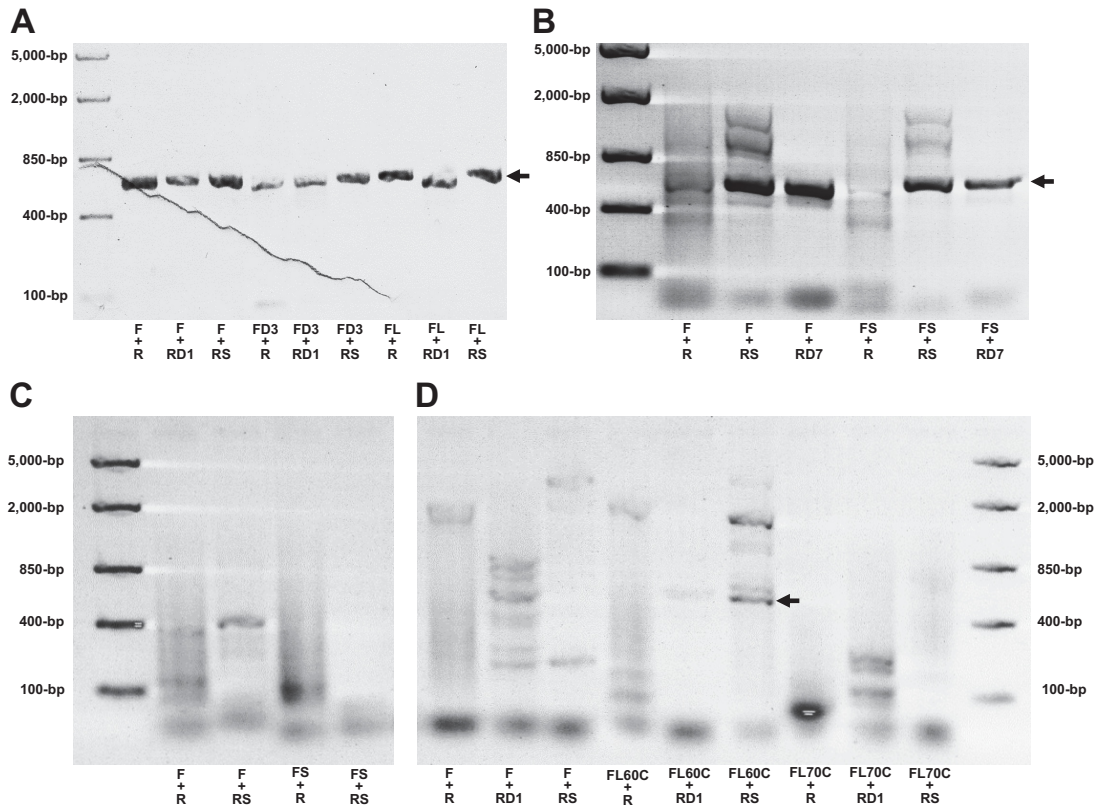


Figure 3-1 –PCR amplification of *mmfR* and *mmyR* orthologues from *S. avermitilis* and *S. sclerotialus*. (**A**) PCR amplification of *avaL1* (657-bp) at 58 °C and (**B**) *avaL2* (621-bp) at 62 °C from the pESAC-13A fragment encoding the azoxy compound gene cluster. (**C**) Attempted amplification of *sclM1* (633-bp) and (**D**) *sclM4* (591-bp) from *S. sclerotialus* genomic DNA extract at 62 °C. Indicated under gels are combinations of primers (listed in Table 2-5). Arrows indicate bands for the target gene of interest (putative in the case of *S. sclerotialus*).

Cloning of the *avaL1, avaL2, sclM1*$_{CO}$ and *sclM4*$_{CO}$ amplicons into the pET151 plasmid and subsequent transformation into *E. coli* TOP10 cells was performed as described in Chapter 2.4. The plasmids were extracted using the Miniprep protocol and insertion of the amplicons

into the pET151 plasmid confirmed by PCR amplification using the $T7_F$ and $T7_R$ primers. The purified pET151 plasmid was then sent for GATC sanger sequencing along with the $T7_F$ and $T7_R$ primers, and the sequenced insert compared against the expected template sequences.

The sequences of the *avaL1*, *avaL2* and *sclM1$_{CO}$* pET151 constructs were identical to the anticipated sequences, but two different *sclM4$_{CO}$* constructs were sequenced from the selected *E. coli* TOP10 colony-forming units 1 and 3 (pET151-*sclM4$_{CO1}$* and pET151-*sclM4$_{CO3}$*), neither of which encoded the expected *sclM4$_{CO}$* sequence. pET151-*sclM4$_{CO1}$* had a 22-bp insertion at the 3' end of the sequence, resulting in a codon frame shift altering the translated termination aa sequence R* (with asterisks depicting the stop codon) to QGRAQIRLLTKPERKLSWLLPPLSNN*. pET151-*sclM4$_{CO3}$* has a single nucleotide deletion in the stop codon that results in an nine amino acid extension of KGELRSGCS*. The work proceeded with the pET151-*sclM4$_{CO3}$*, based on assumed negligible effects of the nine amino acid extension.

3.2.2 Protein Solubility and Purification

After verifying the pET151 insert sequences, the plasmid constructs were transformed and expressed in *E. coli* BL21(DE3) cells. A non-induced culture was used as a control, and samples were separated into soluble and insoluble (or whole-cell) fractions as described in Chapter 2.4. As the proteins were recombinantly expressed, the recombinant molecular weight (*mW$_R$*) of the pET151 constructs differ from that of the native protein sequences because of the addition of the 33 residue N-terminal tag (N-MH_6GKPIPNPLLGLDSTENLYFNGIDPFT-C). To determine the expressed protein concentrations, the molar absorption coefficient at 280 nm was calculated based on the sequence of each recombinant proteins ($\varepsilon 280_R$, Table 3-1).

Table 3-1 – Molecular weight (*mW*) and molar absorption coefficients (*ε*280) of both native and pET151 recombinant (*mW$_R$* & *ε*280$_R$) MmfR/MmyR orthologues from *S. coelicolor* A3(2), *S. avermitilis*, *S. sclerotialus* and *S. venezuelae*.

Protein	*mW* (Da)	*ε*280 (M^{-1} cm^{-1})	*mW$_R$* (Da)	*ε*280$_R$ (M^{-1} cm^{-1})
MmfR	24,052	22,460	27,822	23,950
MmyR	21,883	17,085 (ox.) 16,960 (red.)	25,652	18,575 (ox.) 18,450 (red.)
AvaL1	24,004	11,460	27,774	12,950
AvaL2	22,350	23,950	26,120	25,440
SclM1	23,088	18,450	26,857	19,940
SclM4	21,335	13,980	25,104	15,470
SclM4$_{CO-1}$	24,190	19,480	27,959	20,970
SclM4$_{CO-3}$	22,166	14,105 (ox.) 13,980 (red.)	25,935	15,595 (ox.) 15,470 (red.)
SgnR	22,508	21,430	26,727	22,920
GbnR	21,451	15,470	25,221	16,960

AvaL1, AvaL2, SclM1, SclM4$_{CO-3}$ were initially expressed at 37 °C in 10 mL cultures for 6 hrs. Bands on the SDS-PAGE gels corresponding to the expected recombinant proteins were present for all expressions (Figure 3-2), but only AvaL1 and AvaL2 were within the soluble fraction; SclM1 and SclM4$_{CO-3}$ were not soluble when expressed at 37 °C (not shown). Strong bands were absent in the uninduced cultures, although some leaky-expression is present for AvaL1 expression. These bands were not seen in control cultures which haven't been transformed with pET151 constructs (not shown). The experiment was repeated, with *E. coli* BL21(DE3) cultures induced and incubated overnight at 15 °C. From the overnight cultures SclM1 was found in the soluble fraction, but SclM4$_{CO-3}$ remained insoluble. pET151-*sgnR* (provided by Dr Zhou) was also transformed and expressed, and was similarly only soluble when expressed at 15 °C. Previous work by the Corre group indicates MmfR is also only soluble when expressed at 15 °C, which was confirmed using pET151 constructs encoding MmfR and the five previously-investigated mutants (Y85F, Y85A, Q130E, Y144A, and Y144F) (not shown). Low temperature-dependent solubility appears to be typical for recombinant MmfR-type repressors, for which AvaL1 is the exception.

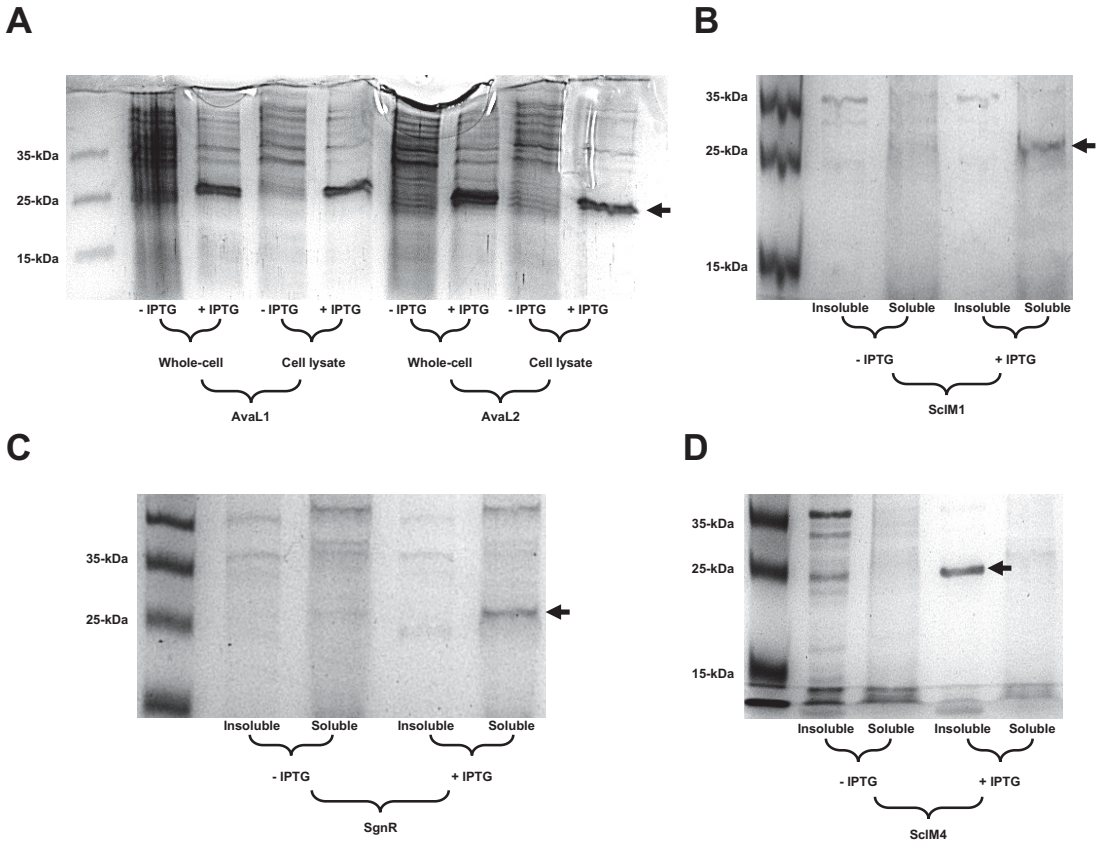
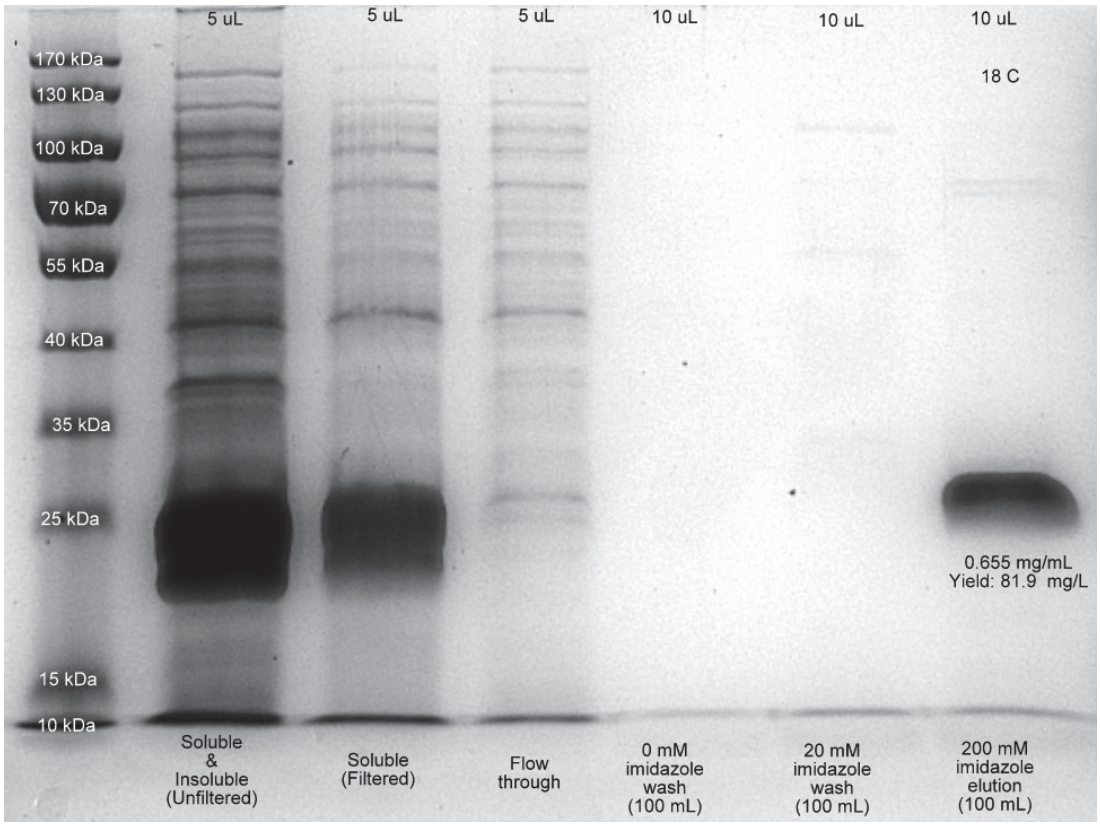


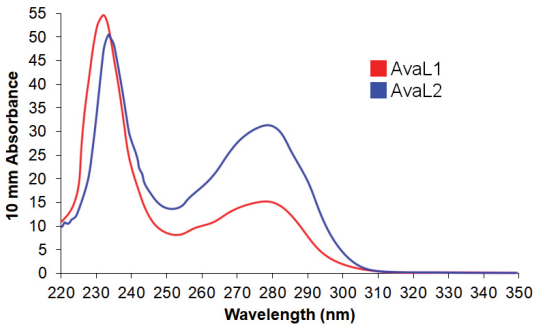
Figure 3-2 – Solubility determination for pET151-recombinant proteins expressed in *E. coli* BL21(DE3). SDS-PAGE gels are shown for (**A**) AvaL1 (27,774 Da) and AvaL2 (26,120 Da) expressed at 37 °C for 6 h, and (**B**) SclM1 (26,857 Da), (**C**) SgnR (26,727 Da) and (**D**) SclM4$_{co-3}$ (25,935 Da) expressed at 15 °C overnight. Arrows indicate bands for expressed protein.

While MmyR solubility was not tested, previous work by the Corre group has indicated that like SclM4$_{CO-3}$, MmyR is also insoluble when expressed at 15 °C (Dr Rea, personal correspondence). The solubility of *S. venezuelae* GbnR has not been determined but based on high sequence identity with MmyR and SclM4 (34.9 %, 38.7 %, and 46.1 % sequence identity with MmyR, AvaL2 and SclM4 respectively) it is probable that GbnR may be similarly insoluble. In which case, like AvaL1, AvaL2 may be an exception to the observed trend of insoluble MmyR-like proteins. After initial validation of recombinant protein solubility, *E. coli* BL21(DE3) cultures were scaled up to 400-800 mL and the protein expressed at 15 °C to 18 °C overnight. The typical yields of expressed AvaL1 and AvaL2 up to ~80 mg L^{-1} meant that Ni^{2+} IMAC purification did not need to be performed often (Figure 3-3A). NanoDrop™ UV spectra of purified AvaL1 and AvaL2 after centrifugal filtration typically had $A_{260/280}$ ratio values close to 0.6, indicating reasonable purity of the solution (Figure 3-3B). The quality control was considered satisfactory for expression of these proteins, as was used for the all protein expressions.

A



B



	AvaL1	AvaL2
$\varepsilon 280_R$ (M^{-1} cm^{-1})	12,950	25,440
[TFR] (mg mL^{-1})	32.04	31.69
[TFR] (mM)	1.15	1.21
$A_{260/280}$	0.66 (0.65 to 0.85)	0.58 (0.55 to 0.65)
Yield (mg L^{-1})	20.03	19.80

Figure 3-3 – Purification examples for pET151-recombinant AvaL1/AvaL2. (**A**) SDS-PAGE gel showing purification of AvaL2 expressed in an 800 mL culture of *E. coli* BL21(DE3) using Ni^{2+} gravity-flow IMAC. AvaL2 purity from this gel is estimated as ~90+ % using ImageJ analysis. Indicated yield is measured post buffer-exchange using a NanoDrop™ 2000. Typical NanoDrop™ 2000 220-350 nm UV spectra for AvaL1 and AvaL2 after purification and concentration in Amicon centrifugal filters. Typical $A_{260/280}$ values ranges are indicated in parenthesis. In the example shown sample concentrations are measured using IMAC Sample Buffer as a blank reference (in other experiments blanks used would be the appropriate buffer for each experiment). Yields are expressed as mg of protein produced per litre of *E. coli* BL21(DE3) in LB liquid media.

3.3 AvaL1/AvaL2/SclM1 activity determination by EMSA

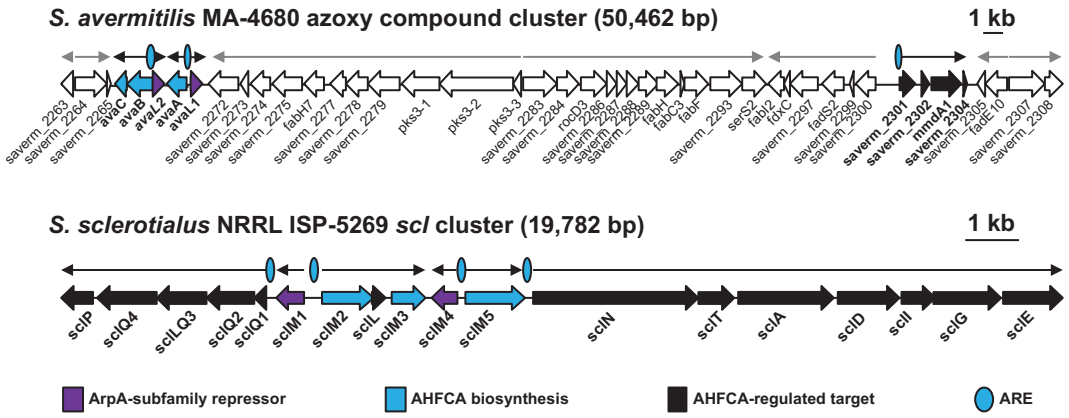
3.3.1 Bioinformatic prediction of azoxy and *scl* gene cluster AREs

To identify AREs in the *S. avermitilis* and *S. sclerotialus* AHFCA-regulated gene clusters, a 24-bp non-palindromic motif based on the *mmfR/mmfLHP*$_{ARE}$, *mmyR*$_{ARE}$ and *mmyB*$_{ARE}$ sequences (Figure 1-9B, Chapter 1) was generated in MEME. The sequences for each AHFCA-regulated gene cluster were parsed through FIMO with a *p* value cut-off of $< 1^{-5}$, which identified three matches in the *S. avermitilis* and four in *S. sclerotialus* gene clusters (Figure 3-4).

In the azoxy compound gene cluster of *S. avermitilis* these sites are found in the divergent promoter regions of *avaL1* and *avaA* (*avaL1/avaA*$_{ARE}$), *avaL2* and the *avaB-avaC* operon (*avaL2/avaBC*$_{ARE}$), and upstream of the SARP-family transcriptional regulator *saverm_2301* (*saverm_2301*$_{ARE}$). Based on these predictions AvaL1/AvaL2 both regulate the biosynthesis of the azoxy compound through regulation of the activator *saverm_2301*, and both AHFCA biosynthetic genes *avaABC* and the regulators *avaL1*/*avaL2* in a manner synonymous with the regulation of *mmyB*, *mmfLHP* and *mmfR*/*mmyR* in the *mmy* gene cluster in *S. coelicolor* A3(2). It is not clear at this point whether the other genes found in the putative operon *saverm_2301-saverm_2305* (as predicted by Dr Poon) would also be coregulated by AvaL1/AvaL2[112], and divergently-positioned *saverm_2300*, which has a distance 1,028-bp between the ARE and the *saverm_2300* start codon, is also not likely regulated. Therefore, this ARE site is labelled only as an ARE for *saverm_2301*.

SclM1 and SclM4 are predicted to regulate the AHFCA regulatory cassette within the *scl* gene cluster in *S. sclerotialus* in a manner similar to *S. venezuelae*, as predicted AREs are found upstream of all putative operons for biosynthetic enzymes[113, 148]. There are two AREs predicted within the AHFCA regulatory cassette; in the divergent promoter region of *sclM1* and the *sclM2-sclL-sclM3* operon (*sclM1/sclM2LM3*$_{ARE}$), and in the divergent promoter region of *sclM4* and *sclM5* (*sclM4/M5*$_{ARE}$). The other two predicted AREs not within the AHFCA regulatory cassette are found upstream of the 5-gene operon encoding *sclQ1-sclQ2-sclQ3-sclQ4-sclP* (*sclQ1-4P*$_{ARE}$) which are responsible for glycolic acid biosynthesis, and upstream of an at the time unannotated ORF which would later be identified as the NRPS *sclN*, the first gene in a putative 7-gene operon comprising *sclN* through *sclE* (*sclN-E*$_{ARE}$) which encode the final SCL biosynthetic enzymes as well as the *sclE* transporter.

A



B

```
avaL1<-n95-AATATACCTtCGCGcAGGTATATT-n72 ->avaA
avaL2<-n19-AagATaCGTTCgtGAACGaATtcT-n25 ->avaB
          …AaAATaCCTTctcaAAGGaATTaT-n178->saverm_2301

sclM1<-n114-AATATACCTtcccatAGGTATATT-n63 ->sclM2
sclM4<-n34 -AAgaAaCttCCaaGGcgGaTctTT-n116->sclM5
          …tTgAAACCTTCGCGAAGGTTTgAt-n111->sclQ1
          …tgaATaCCTcCGCGaAGGaATgat-n40 ->sclN
```

Figure 3-4 – Bioinformatically predicted MmfR-type AREs in the AHFCA-regulated gene clusters of *S. avermitilis* and *S. sclerotialus*. (**A**) Genetic organisation of the *S. avermitilis* and *S. sclerotialus* gene clusters annotated with predicted AREs. (**B**) sequences of predicted ARE sites. Half-dyads around the dyad centre are coloured **red** and **blue** respectively, with palindromic bases indicated in upper case. n indicates the number of nucleotides between the ARE and the start codon.

3.3.2 Binding of AvaL1 to *S. avermitilis* AREs

EMSA protocols and preparation of DNA for assaying protein-DNA interactions are detailed in Chapter 2.4.11, and the oligonucleotide sequences used are shown in Table 2-7, Chapter 2.2. The binding of AvaL1 to *avaL1/avaA*$_{ARE}$, *saverm_2301*$_{ARE}$ and *avaL2/avaBC*$_{ARE}$ was assayed over a molar ratio range of 1:(0.25-8) DNA:AvaL1, with the DNA concentration fixed at 200 nM (Figure 3-5). AvaL1 was capable of binding to all three predicted ARE sequences in the *S. avermitilis* azoxy compound gene cluster. The free DNA shifted to the top of the gel at the lowest assayed AvaL1 concentration, increasing until the DNA was visibly saturated at a ratio of ~1:4. Single free DNA and complex bands are observed in the *saverm_2301*$_{ARE}$ gel, indicating the intended preparation of the hairpin dsDNA probe. There is no indication of different stoichiometric complexes of AvaL1 and DNA, but this may be due to insufficient separation of bands due to poor mobility of the complexes through the gel. This also appears to be the case for both the AvaL1 homologue ArpA and CprB, for which gel shift assays also do not indicate the formation of multiple complexes due to poor migration of the complexes through the gel[69, 152, 204].

Two high-mobility bands in the free DNA region of the *avaL1/avaA$_{ARE}$* and *avaL2/avaBC$_{ARE}$* gels indicate incomplete preparation of the self-annealed oligonucleotide. As the two free DNA bands correlate with the two observed complex bands, the second band is not likely to be unannealed ssDNA based on the assumption that AvaL1 only recognises dsDNA specifically. The band intensity of the two complex bands are similar; therefore, the amount of bound DNA is similar, being more intense than would be expected for binding of AvaL1 to ssDNA. Therefore, these bands most likely belong to two dsDNA sequences. These two bands are rationalised as a cooccurrence of both the hairpin dsDNA probe and dsDNA with two intact ARE sites which has either inadequately denatured or reannealed from two ssDNA sequences. The less mobile band is the larger, dsDNA, and the other the hairpin dsDNA that migrates further through the gel. Additionally, the stock dsDNA probe will have twice the number of binding sites per probe, resulting in further retardation upon AvaL1 binding. Faint bands are also observed near the top of the *avaL1/avaA$_{ARE}$* and *avaL2/avaBC$_{ARE}$* gels, between the free DNA and complexed DNA bands which, based on shifting of these bands during AvaL1 binding, suggests that these are also free DNA bands. Based on poor migration of ssDNA through gel compared to dsDNA, this is possibly unannealed ssDNA which aggregates non-specifically with AvaL1 when AvaL1 was added to the samples.

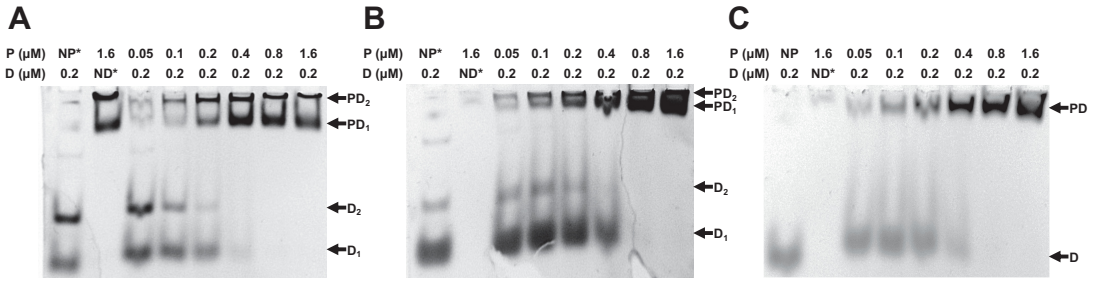


Figure 3-5 – Gel shift assays showing AvaL1 binding to the three putative *S. avermitilis* AREs within the azoxy compound gene cluster. Shown are gels for AvaL1 binding to (**A**) *avaL1/avaA$_{ARE}$*, (**B**) *avaL2/avaBC$_{ARE}$* and (**C**) *saverm_2301$_{ARE}$*. D is DNA, P is protein, PD is protein:DNA complex, NP is no protein control, and ND is no DNA control. All gels are stained with the GelRed intercalator and visualised under a transilluminator. Asterisk indicates leakage from adjacent cell during pipetting. Multiple D and PD bands (numbered) are observed for binding which are attributed to inadequate preparation of the dsDNA hairpin oligonucleotides. A smaller possible band (between D_2 and PD_1 (for **A** and **B**) may be attributed to ssDNA. All three gels shift assays were performed concurrently in the same Mini-PROTEAN® Tetra Cell.

3.3.3 Binding of AvaL1 to consensus AREs

Two additional probes were designed base on MEME analysis of the AREs found in the *S. avermitilis* azoxy compound gene cluster. *SAV_CS_V1$_{ARE}$* was based on the nonpalindromic consensus of the most frequent nucleotides, 5'-AA_ATaCCTTCG|CGAAGGaATT_T-3', with underscored positions composed of nucleotides A and T in the *SAV_CS_V1$_{ARE}$* sequence (5'-AA**A**ATaCCTTCG|CGAAGGaATT**T**T-3'). The *SAV_Pal$_{ARE}$* probe is the palindromic consensus of nucleotides for the three *S. avermitilis* AREs, which has a single nucleotide

difference from *SAV_CS_V1$_{ARE}$* (5'-AAAATACCTTCG|CGAAGG**T**ATTTT-3'). AvaL1 bound to these two additional consensus sequences, but with slightly less shifting of DNA observed at higher ratios of protein-DNA compared with the binding of AvaL1 to the endogenous AREs (Figure 3-6). Better preparation of the self-annealing oligonucleotide resulted in single bands for *SAV_CS_V1$_{ARE}$*, which only produced a single shifted band in the complex region, although a weak second dsDNA band is seen in the *SAV_Pal$_{ARE}$* sequence. Again, there is no indication of complex stoichiometries in the *SAV_CS_V1$_{ARE}$* gel, and a possible second band in the *SAV_Pal$_{ARE}$* could be due to the second free DNA band.

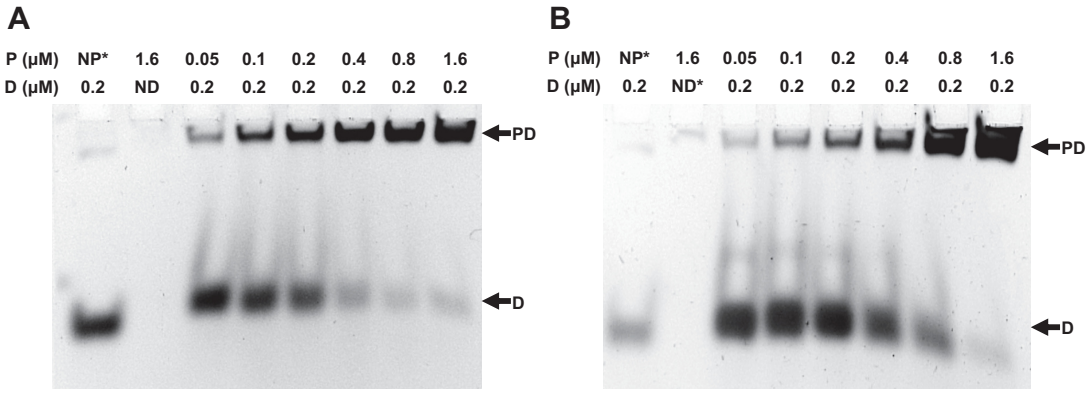


Figure 3-6 – Gel shift assays showing AvaL1 binding to consensus AREs based on sequences identified in the S. avermitilis azoxy compound gene cluster. Shown are gels for AvaL1 binding to (**A**) *SAV_CS_V1$_{ARE}$* and (**B**) *SAV_PAL$_{ARE}$*. D is DNA, P is protein, PD is protein:DNA complex, NP is no protein control, and ND is no DNA control. Asterisk indicates leakage from adjacent cell during pipetting. Both gels shift assays were performed concurrently in the same Mini-PROTEAN® Tetra Cell.

3.3.4 Binding of AvaL2 to *S. avermitilis* AREs

To test the ability of AvaL2 to bind to the *S. avermitilis* AREs, the molar ratio range was expanded to 1:(0.5-16) of DNA:AvaL2, with a FastRuler™ LR DNA ladder added to confirm the size of the probe. The gels (Figure 3-7) indicate that the AvaL2 repressor showed poor binding to the AREs recognised by AvaL1, showing only non-specific binding. The assay was repeated with an expanded molar range to a 1:(16-512) molar ratio, but again no significant binding was observed; only an increase in non-specific binding. Non-specific binding observed here is proposed to be due to the functional homology of AvaL2 as an ArpA-subfamily repressor, but it is possible this could have been misfolded aggregate protein (which could be removed in future gel shifts by size-exclusion chromatography (SEC)). A single free DNA band was observed, and the size of the band was smaller than the lowest 50-bp marker, which is near where the non-hairpin dsDNA probe would be expected to be (59-bp). This is consistent with the hairpin dsDNA probe, as ssDNA would not be expected to migrate that far through the gel.

The lack of AvaL2 binding to AvaL1-recognised AREs correlates with the lack of MmyR binding to the *S. coelicolor* A3(2) AREs performed *in vivo* by Dr Styles[149]. These AREs therefore are likely not the true target sites for this repressor in its native configuration. The possibility that another component may be required for binding to occur cannot be excluded, but if AvaL2 truly represses genes in the biosynthetic gene cluster, it is likely through binding to other sites or indirectly by controlling a secondary repressor. It is possible that the AvaL2 AREs could be within the same promoter regions as the AvaL1 target sites; assaying these regions in future experiments would validate the existence of these sites. It is not yet possible to tell if AvaL2, like ScbR2, is not auto-regulatory in nature, but the nature of regulation by MmfR in the *mmy* gene cluster in *S. coelicolor* A3(2) suggest AvaL2 expression may be repressed primarily through the binding of AvaL1 to the *avaL2/avaBC*$_{ARE}$. It is possible that the other pseudo-AHFCA receptors, GbnR and SclM4, behave in a similar way with recognition of their target sites.

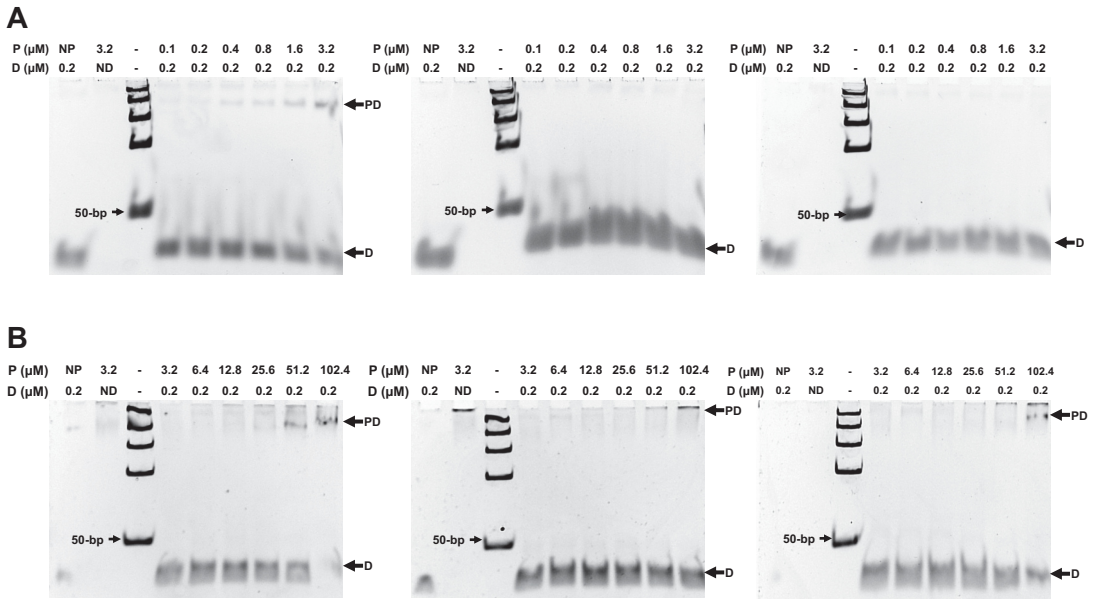


Figure 3-7 – Gel shift assays showing poor binding of AvaL2 to *S. avermitilis* AREs within the azoxy compound gene cluster. Shown from left to right is AvaL2 being assayed for binding to *avaL1/avaA*$_{ARE}$, *avaL2/avaBC*$_{ARE}$ and *saverm_2301*$_{ARE}$, over DNA:AvaL2 molar ratios of (**A**) 1:(0.5-16) and (**B**) 1:(16-512). D is DNA, P is protein, PD is protein:DNA complex, NP is no protein control, and ND is no DNA control. The ladder present is a FastRuler™ Low Range DNA ladder. All three gels shift assays for each set of molar ratios were performed concurrently in the same Mini-PROTEAN® Tetra Cell.

3.3.5 Binding of SclM1 to *S. sclerotialus* AREs

The SclM1 repressor was assayed against the predicted AREs in the *scl* gene cluster at an expanded molar ratio of 1:(0.25-16) DNA:protein (Figure 3-8). SclM1 induced DNA band shifts at similar rates to AvaL1 at all four target sites, with DNA almost saturated at a 1:16 molar ratio. Disparity in the palindromacy between the *sclM4/sclM5*$_{ARE}$ sequence and the other ARE sequences suggests that the dependence of MmfR orthologue recognition for palindromic sequences may not be as dependent on palindromacy as previously thought. The SclM1:DNA complex bands at the top again show poor migration through the gel. However, the appearance of two possible stoichiometric configurations of SclM1 (corresponding to a single dimer and a pair of dimers bound to DNA) are observed in the binding of SclM1 to *sclN-E*$_{ARE}$ and *sclQ1-4P*$_{ARE}$, where there is a shift in position of complexed DNA between 1.6 µM and 3.2 µM SclM1. There are only single bands of free DNA observed in the sample and negative control wells, so these additional complexed bands are likely not as a result of the shift of multiple free DNA bands. The observed shifts in complexed DNA at high concentrations may be an indicator that shifting to a complex comprised of a pair of dimers is either concentration dependent and/or dependent on saturation of DNA. Confirmation of binding stoichiometry could be performed by SPR or analytical ultracentrifugation.

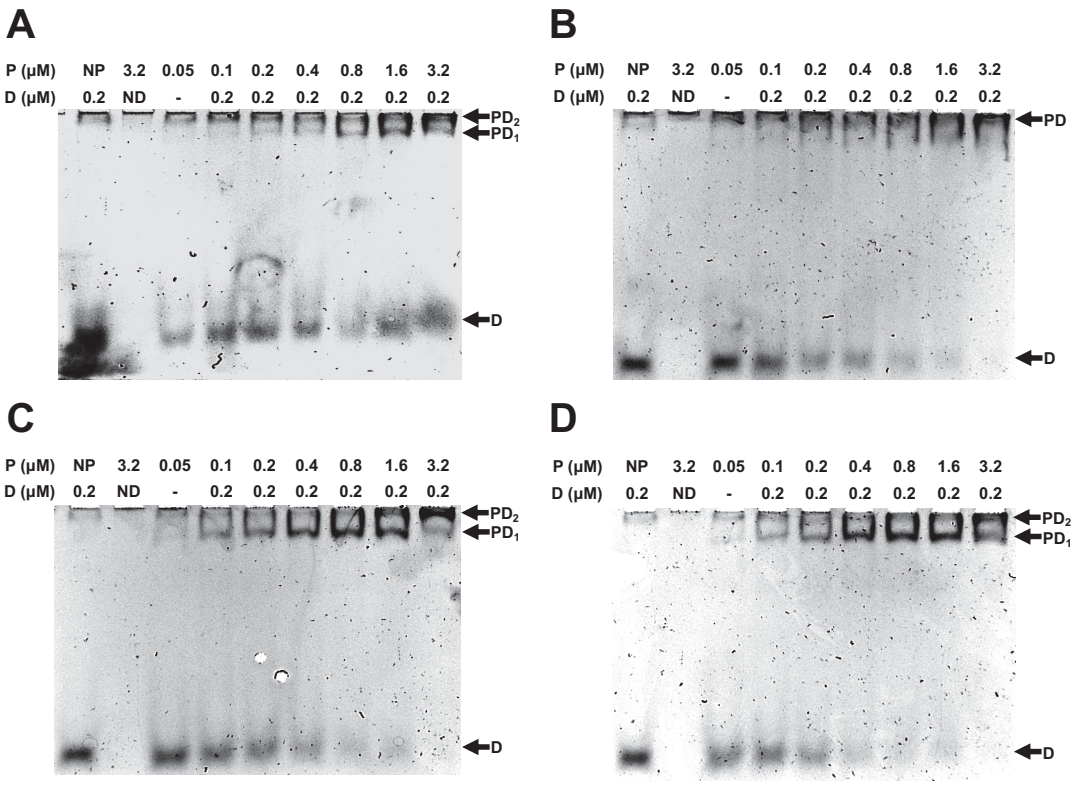


Figure 3-8 – Gel shift assays showing SclM1 binding to the four putative *S. sclerotialus* AREs within the *scl* gene cluster. AREs assayed were (**A**) *sclM1/sclM2LM3*$_{ARE}$, (**B**) *sclM4/sclM5*$_{ARE}$, (**C**) *sclN-E*$_{ARE}$ and (**D**) *sclQ1-4P*$_{ARE}$. D is DNA, P is protein, PD is protein:DNA complex, NP is no protein control, and ND is no DNA control. All four gels shift assays were performed concurrently in the same Mini-PROTEAN® Tetra Cell.

3.3.6 Binding of AvaL1 to exogenous *S. sclerotialus* AREs

Dr Zhou previously reported that the MmfR orthologue SgnR in *S. venezuelae* was capable of binding to the *mmfR/mmfLHP* intergenic region in *S. coelicolor* A3(2)[148], presumably at the same ARE bound by MmfR. Gel shift assays were performed for AvaL1 binding to the AREs identified in *S. sclerotialus*. A positive control of AvaL1 binding to *S. avermitilis* AREs was not included because of high sequence similarity between *S. avermitilis* and *S. sclerotialus* ARE sequences, with *avaL1/avaA*$_{ARE}$ and *sclM1/sclM2LM3*$_{ARE}$ being the most similar (5'-AATATACCT**r**CG|**s**GAAGGTATATT-3', where r = A/G and s = G/C). AvaL1 was also able to both recognise and bind strongly to the four AREs (Figure 3-9) with similar shifts to those observed in the previous gel shift assays. Band shifts consistent with the formation of two different sized complexes as observed for SclM1 also appear to be present here, but this may be because of residual free non-hairpin dsDNA, which can be faintly seen in some lanes (but not the negative controls). These results combined with Dr Zhou's results suggest that the four experimentally investigated MmfR orthologues are capable of recognising AREs in other *Streptomyces* species, and thus share a common mechanism of DNA recognition, although the binding of MmfR and SclM1 to non-endogenous sequences remains to be confirmed.

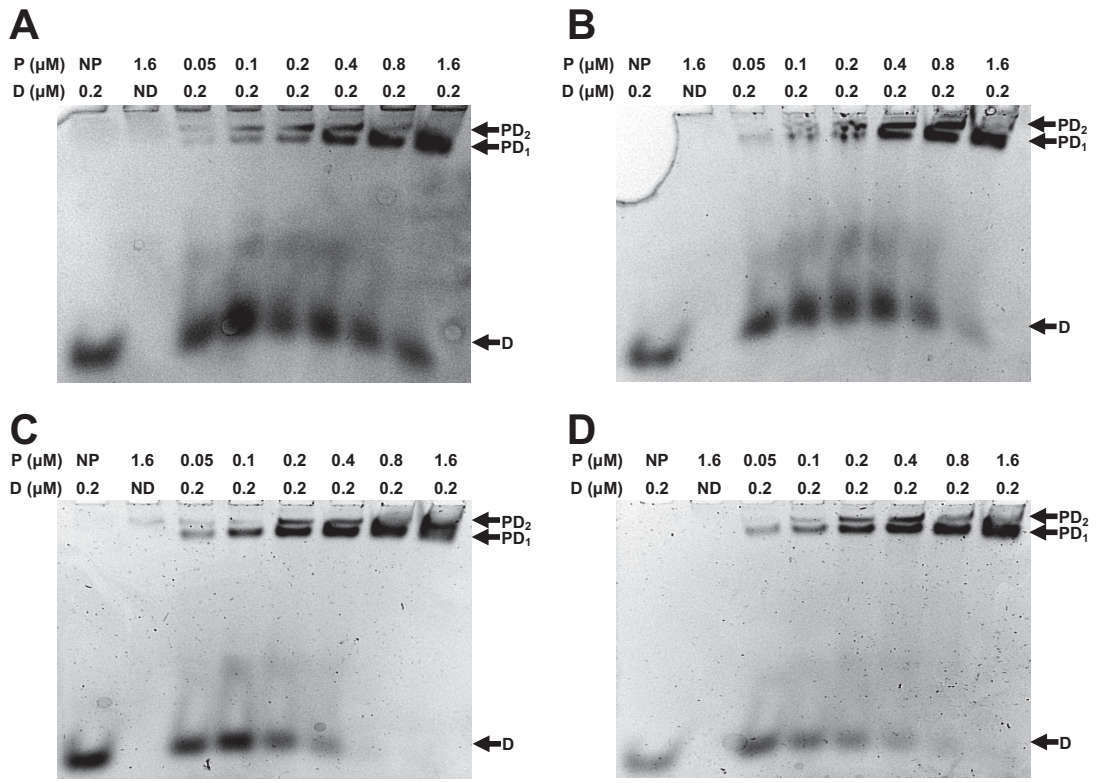


Figure 3-9 – Gel shift assays showing AvaL1 binding to the four *S. sclerotialus* AREs within the *scl* gene cluster. AREs assayed were (**A**) *sclM1/sclM2LM3*$_{ARE}$, (**B**) *sclM4/sclM5*$_{ARE}$, (**C**) *sclN-E*$_{ARE}$ and (**D**) *sclQ1-4P*$_{ARE}$. D is DNA, P is protein, PD is protein:DNA complex, NP is no protein control, and ND is no DNA control. All four gels shift assays were performed concurrently in the same Mini-PROTEAN® Tetra Cell.

3.3.7 AvaL1 and SclM1 derepression by AHFCA6

To assay the ability of AHFCAs to release AvaL1 and SclM1 from DNA, AvaL1:*saverm_2301*$_{ARE}$ and SclM1:*sclQ1-4P*$_{ARE}$ were pre-bound at a molar ratio of 1:32 for 15 min before being mixed with AHFCA(1-4) at a molar ratio of 1:(975-3,900) AvaL1/SclM1:AHFCA (not shown). However, these initial gel shift assays failed to indicate the release of the putative AHFCA receptors from the AREs due to poor DNA-binding by the protein samples. For repeat experiments the DNA:protein:ligand ratios were altered to better observe release of protein from DNA by reducing the concentrations of TFR bound to DNA, and the experiment was repeated only for AvaL1. AvaL1 was pre-bound to *saverm_2301*$_{ARE}$ at a molar ratio of 1:4 DNA:protein, and AHFCA6 was added at a molar ratio of 1:10 protein:ligand; a final ratio of 1:4:40 *saverm_2301*$_{ARE}$:AvaL1:AHFCA6. The complex partially dissociated in the presence of AHFCA6 (Figure 3-10), indicating both that AvaL1 is an AHFCA receptor, and this it is specifically receptive to AHFCA6.

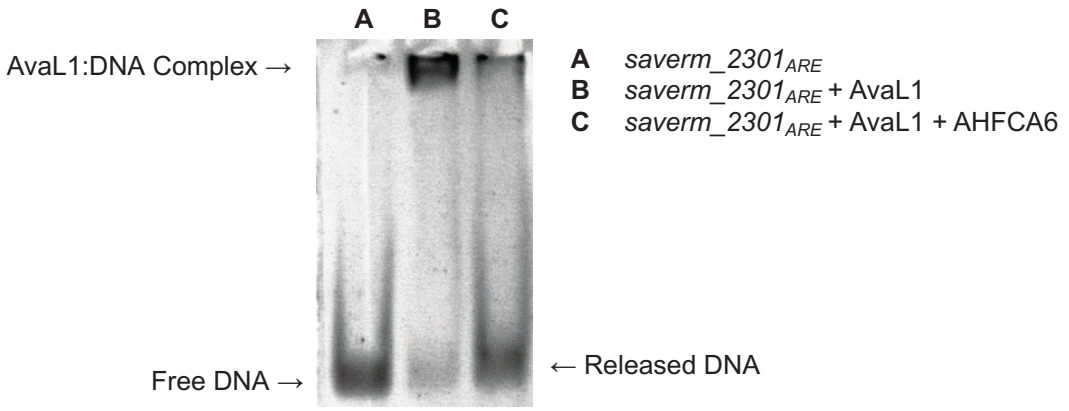
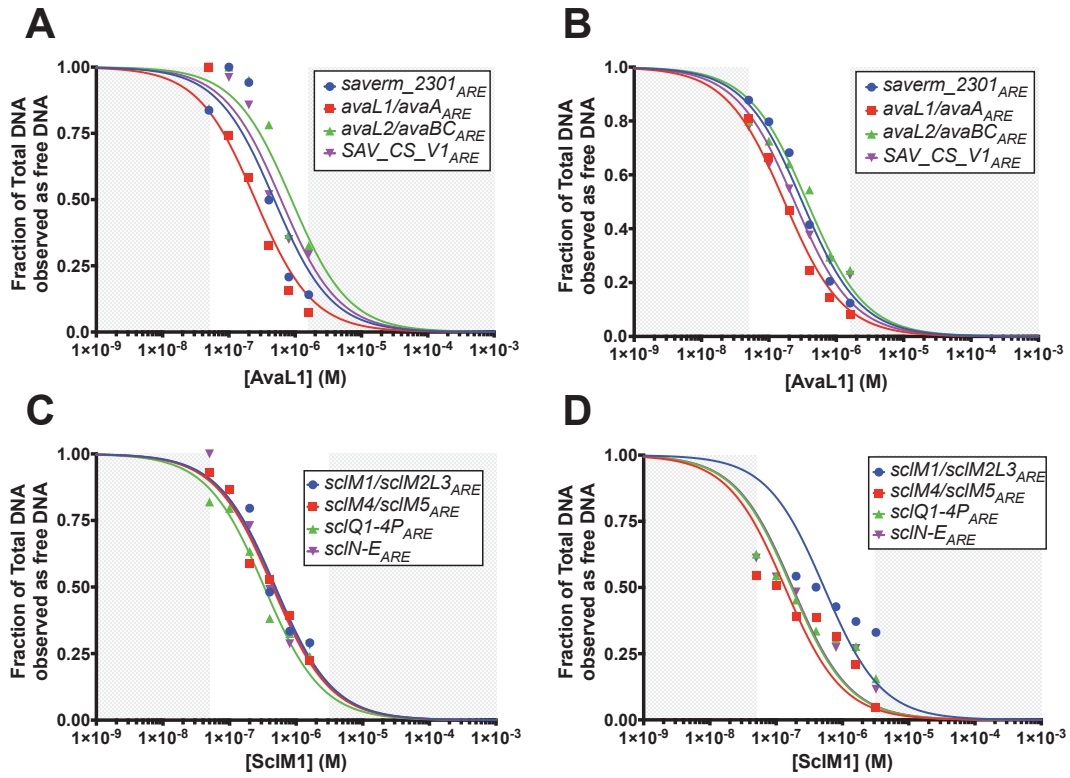


Figure 3-10 – EMSA gel showing the release of AvaL1 from *saverm_2301*$_{ARE}$ by AHFCA6.

3.3.8 Semiquantitative analysis of AvaL1/SclM1 ARE binding affinity

Using the gel images of the gel shift assays shown prior, the affinity of the repressors for their DNA target sequences could be estimated either by measuring the shift in intensities of free DNA bands from the bottom of the gel (with an IC_{50} curve fitted) or the shift of DNA to the protein-DNA complex at the top of the gel (fitted with a K_D curve) using ImageJ. Fitting of the curves were constrained to fractions of 0 to 1 of relative DNA band intensity, with the highest intensity band taken within the group taken as 1 and the intensity of an empty lane as 0. Assuming that all DNA could be accounted for in the sum of the free DNA and complexed DNA fractions within a single lane, IC_{50} was also derived with the data normalised such that the sum of band intensities for each gel lane = 100 % of accountable DNA (IC_{50}^{N}). The quality

of free DNA band intensities from the gels for AvaL1 binding to the *S. sclerotialus* AREs as well as SAV_PAL_{ARE} meant that fitting of IC_{50} curves could not be fitted to the derived intensities. The same goes for binding of AvaL2 to the *S. avermitilis* AREs. However, K_D estimation based on complex formation did generate some meaningful data from these gels.



E

TFR:ARE complex	IC_{50} (nM)	IC_{50}^N (nM)	Statistical Tests
AvaL1:*saverm_2301*$_{ARE}$	482 ± 154	312 ± 33	**AvaL1 IC_{50} vs IC_{50}^N**
AvaL1:*avaL1/avaA*$_{ARE}$	250 ± 45	170 ± 12	$t(3) = 2.91$, $p = 0.06$
AvaL1:*avaL2/avaBC*$_{ARE}$	878 ± 350	368 ± 43	
AvaL1:*SAV_CS_V1*$_{ARE}$	609 ± 152	240 ± 29	**SclM1 IC_{50} vs IC_{50}^N**
			$t(3) = 2.25$, $p = 0.11$
Mean	**555 ± 414**	**273 ± 63**	
Model Fit R^2	**0.80 ± 0.08**	**0.97 ± 0.01**	**AvaL1 IC_{50} vs SclM1 IC_{50}**
			$t(6) = 0.90$, $p = 0.19$
SclM1:*sclM1/sclM2n3*$_{ARE}$	489 ± 74	529 ± 156	
SclM1:*sclM4/sclM5*$_{ARE}$	443 ± 47	136 ± 36	**AvaL1 IC_{50}^N vs SclM1 IC_{50}^N**
SclM1:*sclQ1-4P*$_{ARE}$	332 ± 34	170 ± 41	$t(6) = 0.40$, $p = 0.85$
SclM1:*sclN-E*$_{ARE}$	466 ± 63	177 ± 42	
Mean	**433 ± 113**	**253 ± 171**	
Model Fit R^2	**0.96 ± 0.01**	**0.79 ± 0.06**	

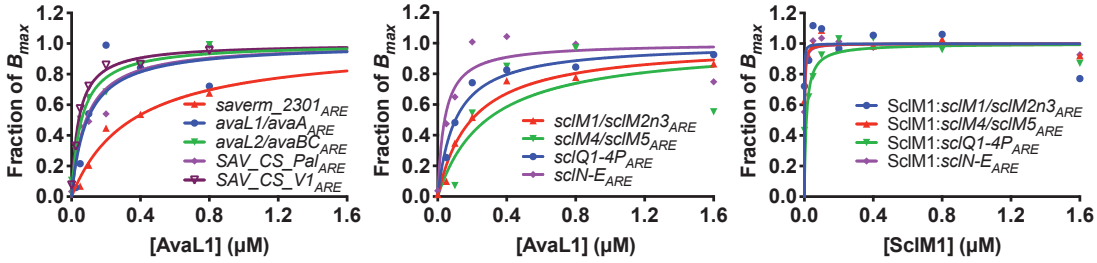
Figure 3-11 – Semiquantitative determination of affinity for AvaL1 and SclM1 binding to endogenous and consensus AREs using IC_{50} curves. (**A**) Constrained IC_{50} curves fitted to free-DNA band intensities for AvaL1 binding to *S. avermitilis* endogenous AREs plus $SAV\_CS\_V1_{ARE}$ with an Ordinary Least Squares regression. (**B**) As for **A** with same free DNA band intensities normalised against complexed DNA band intensities (for IC_{50}^N determination). (**C**) As for **A** for SclM1 binding to *S. sclerotialus* AREs, and (**D**) data from **C** normalised as for **B**. (**E**) Table of derived IC_{50} and IC_{50}^N values for binding of AvaL1 and SclM1 to target AREs. Model Fit R^2 are given as mean values for fitting of individual IC_{50} curves. The variance between derived IC_{50} and IC_{50}^N values for all AREs (AvaL1 and SclM1) are homoscedastic ($W(3, 12) = 2.02$, $p = 0.16$). Unpaired Student's t test statistics are given for the comparison of the binding for AvaL1 vs SclM1 for both IC_{50} and IC_{50}^N (variance was homoscedastic; $W(1, 6) = 3.41$, $p = 0.11$), in addition to paired Student's t test statistics for IC_{50} vs IC_{50}^N for both AvaL1 and SclM1.

For the binding of AvaL1 to *S. avermitilis* AREs (Figure 3-11A and Figure 3-11B) the IC_{50}^{N} curve had the stronger fit with the data, and inversely the IC_{50}^{N} fitting was the poorer fit for the binding of the SclM1 to *S. sclerotialus* AREs (Figure 3-11C and Figure 3-11D). AvaL1 bound to target AREs at a sub-μM affinity, with mean IC_{50} and IC_{50}^{N} values of 555 nM ± 414 nM and 273 nM ± 63 nM respectively. For SclM1 binding to *S. sclerotialus* AREs the mean IC_{50} and IC_{50}^{N} were similarly 433 nM ± 113 nM and 253 nM ± 171 nM. There was no significant difference between the mean IC_{50} and IC_{50}^{N} values calculated for either AvaL1 (paired Student's $t(3) = 2.91$, $p = 0.06$) or SclM1 (paired Student's $t(3) = 2.25$, $p = 0.11$), and there was no significant difference in mean IC_{50}/IC_{50}^{N} values for AREs bound by either AvaL1 or SclM1 (for IC_{50} values unpaired Student's $t(6) = 0.90$, $p = 0.40$, and for IC_{50}^{N} $t(6) = 0.19$, $p = 0.85$).

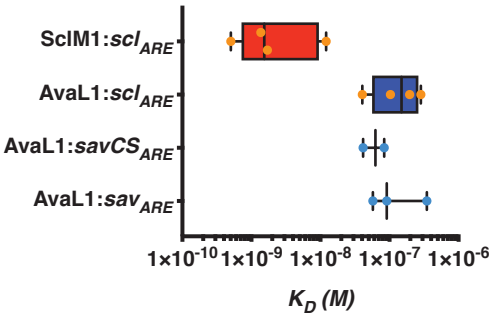
The precision of the determined K_D values by comparison with IC_{50} and IC_{50}^{N} values (especially for SclM1) is much poorer. (Figure 3-12). The variance was homoscedastic between all IC_{50}, $IC50_N$ and K_D values ($W(2, 26) = 0.50$, $p = 0.61$), and only the mean IC_{50} and K_D values for ARE binding by SclM1 being significantly different (paired Student's $t(3) = 11.88$, $p < 0.01$). Within the K_D values the binding of AvaL1 to endogenous, exogenous and consensus sequences all have equal mean K_D values (ANOVA $F(2, 6) = 0.52$, $p = 0.62$), which are significantly different from the values for SclM1 binding to endogenous AREs (unpaired Student's $t(11) = 2.33$, $p = 0.04$).

The disparity in AvaL1 and SclM1 K_D values is most likely due to the poor image quality of the SclM1:DNA complex bands. Factoring in the IC_{50} and IC_{50}^{N} data, AvaL1 likely shares a similar affinity for DNA as SclM1 and judging by the similarities between the binding of both endogenous and exogenous AREs by AvaL1, both repressors shares a common mechanism that allows AvaL1 to bind to the same sites as SclM1 would *in situ* at a similar affinity. The range of predicted K_D/IC_{50} values is of greater relevance than distinguishing the differences in affinity between sequences as the precision is not strong enough to tell these apart (as supported in later SPR experiments). The range of values can therefore be used as points of comparison using SPR and future experiments.

A



B



Total Mean K_D = 97 ± 690

C

ARE	AvaL1 K_D (nM)	SclM1 K_D (nM)
saverm_2301$_{ARE}$	348 ± 192	
avaL1/avaA$_{ARE}$	91 ± 67	
avaL2/avaBC$_{ARE}$	57 ± 15	
SAV_Pal$_{ARE}$	84 ± 33	
SAV_CS_V1$_{ARE}$	41 ± 7	
Mean	**124 ± 207**	
Model Fit R^2	**0.92 ± 0.06**	
sclM1/sclM2n3$_{ARE}$	196 ± 47	1 ± 9
sclM4/sclM5$_{ARE}$	284 ± 656	1 ± 8
sclQ1-4P$_{ARE}$	103 ± 23	12 ± 10
sclN-E$_{ARE}$	40 ± 24	2 ± 7
Mean	**156 ± 659**	**4 ± 17**
Model Fit R^2	**0.83 ± 0.22**	**-1.37 ± 1.39**

D

Statistical Test	Test Results
IC_{50} **vs** IC_{50}^N **vs** K_D	Levene's $W(2, 26) = 0.50$, $p = 0.61$)
AvaL1 – IC_{50} **vs** K_D	Unpaired Student's $t(3) = 2.52$, $p = 0.09$
AvaL1 – IC_{50}^N **vs** K_D	Unpaired Student's $t(3) = 1.84$, $p = 0.16$
SclM1 – IC_{50} **vs** K_D	Paired Student's $t(3) = 11.88$, $p < 0.01$
SclM1 – IC_{50}^N **vs** K_D	Paired Student's $t(3) = 2.69$, $p = 0.07$
K_D (AvaL1:*sav/savCS$_{ARE}$* **vs** AvaL1:*scl$_{ARE}$* **vs** SclM1:*scl$_{ARE}$*)	Levene's $W(2, 10) = 3.20$, $p = 0.08$
K_D (AvaL1:*sav$_{ARE}$* **vs** AvaL1:*savCS$_{ARE}$* **vs** AvaL1:*scl$_{ARE}$*)	ANOVA $F(2, 6) = 0.52$, $p = 0.62$
K_D (AvaL1:*sav/savCS/scl$_{ARE}$* **vs** SclM1:*scl$_{ARE}$*)	Unpaired Student's $t(11) = 2.33$, $p = 0.04$

Figure 3-12 – Semiquantitative determination of affinity for AvaL1 and SclM1 binding to endogenous, consensus and exogenous AREs using saturation curves for complex bands. (**A**) K_D Saturation curves fitted with an Ordinary Least Squares regression for the binding of both repressors to each ARE as indicated. *sav* = *S. avermitilis*, *scl* = *S. sclerotialus* and *CS* = consensus sequence. (**B**) Box and Whisker plots for mean K_D for sequences; boxes are coloured according to repressor, while data points are coloured according to ARE origin. (**C**) Table of the derived K_D values. Model Fit R^2 are given as mean values for fitting to individual saturation curves. (**D**) Table of performed statistical tests (using data from Figure 3-11).

3.4 Crystallography of AvaL1, AvaL2 and SgnR

3.4.1 Initial crystallisation trials

AvaL1, AvaL2 and SgnR were initially screened with Morpheus®, HCSG-plus™, PACT premier™ and Wizard Classic 1 and 2 kits at 26, 22 and 31 mg mL^{-1} respectively in both the ligand-free and ligand-bound configurations on the Mosquito® LCP ((Figure 8-2 and Figure 8-3, Appendix B). 500 mM NaCl IMAC sample buffers without glycerol were used in all crystallisation trials. Of these screening kits both Morpheus® and Wizard Classic screens produced initial hits of interest (Table 3-2). The Morpheus® screen hits frequently occurred in crystallant conditions with carboxylic acid (row G) or amino acid (row H) additive mixes in either Buffer System 1 (1 M imidazole/H_2O.MES, pH 6.5) or Buffer System 2 (1M NaHEPES/MOPS, pH 7.5). In addition, one hit was identified with monosaccharide additives (F) in Buffer system 3 (1 M Tris (base)/BICINE, pH 8.5). Precipitant mix 3 (40 % v/v Glycerol, 20 % w/v PEG 4000) had five hits, compared to three hits with precipitant mix 1 (40 % v/v PEG 500 MME, 20 % w/v PEG 20,000) and three hits in precipitant mix 2 (40 % v/v Ethylene glycol; 20 % w/v PEG8000). The Wizard screen had only three crystallant hits; C9 (2 M ammonium sulfate, 100 mM CAPS/sodium hydroxide pH 10.5, 200 mM lithium sulfate), E3 (20 % v/v PEG 8,000, 100 mM Tris (base)/HCl pH 6.0, 200 mM magnesium chloride) and E11 (10 % v/v 2-propanol, 100 mM sodium cacodylate/HCl pH 6.5, 200 mM lithium sulfate), with a not too dissimilar set of morphologies.

Table 3-2 – Crystallant conditions which produced positive hits for AvaL1, AvaL2 and SgnR crystallisation during initial crystallisation trials. All trials were performed on a Mosquito® LCP, with AvaL1, AvaL2 and SgnR concentrations of 26, 22 and 31 mg mL^{-1} respectively. Protein:ligand was mixed at a molar ratio of 1:2 protein:ligand. Images can be found in Figure 8-2 and Figure 8-3, Appendix B.

	Morpheus®										Wizard Classic 1 and 2		
	F11	G2	G3	G5	G7	H1	H3	H5	H6	H7	C9	E3	E11
AvaL2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
AvaL1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
AvaL1:AHFCA1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
AvaL1:AHFCA3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	?	✓	✓
AvaL1:SCB1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SgnR	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
SgnR:AHFCA5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

These hits primarily produced either phase separation or a dense anomalous precipitated aggregate with varying morphology, and commonly gelatinous in density. Some conditions produced aggregates that formed along branching interfaces, either around or protruding from the circumference of the droplet, while some structures were formed from within the amorphous precipitate. Although not identical for all proteins, the crystallant conditions screened had similar morphologies for each protein screened.

Co-crystallisation of the protein with ligands resulted in very little change in morphology compared with the ligand-free protein screens. This was expected, given that the crystallisation conditions for both MmfR and MmfR:AHFCA2 were identical. Generally, many of the precipitant morphologies were the same; for AvaL1:AHFCA1 the formation of regulator structures in some droplets was lost, and slightly diminished in others. For AvaL1:AHFCA5 precipitants were more gelatinous than the ligand-free precipitants. AvaL1:SCB1 precipitants were more amorphous, while for SgnR:AHFCA5 the formation of regular precipitant structures was sometimes improved, but otherwise very similar to ligand-free SgnR. Any changes that occurred may be batch/droplet variance, or due to the presence of free ligand or DMSO solvent affecting the composition of the droplet.

All hits from the Morpheus and Wizard screen for AvaL1 were scaled up onto manual 2:2 µL 24-well screens, but only brown denatured precipitate was formed (not shown). Wizard C9, which initially produced only densely denatured protein aggregate, later had apparent needle formations after several weeks (Figure 8-4, Appendix B). Whether these crystals were protein or salt crystals is not clear. Both AvaL1 and AvaL2 were also screened in the same crystallant condition as for MmfR in a 24-well optimisation grid for 0.20 M to 0.25 M magnesium formate and 10 % to 16 % PEG 3,000 (closest available substitute for PEG 3,350). AvaL1 produced only clear drops, while the AvaL2 screen formed only phase separation (not shown).

The Morpheus® kit was rescreened for AvaL1 at higher drop ratios (300:300 nL and 300:100 nL) on the Mosquito® LCP instrument (Figure 8-4, Appendix B). The increase in droplet volumes from 100:100 nL produced less amorphous precipitate than was observed previously. There was little difference between the 100:300 nL and 300:300 nL drop volumes. Precipitates had either a gel-like consistency or were covered by skin-like formations, possibly layers of aggregated or denatured protein/PEG. The hits were not identical to those found in the initial screen; most hits were of buffer system 3/precipitant mix 1 for all additives. Additional hits were found for amino acid additive conditions H1 (Buffer system 1, precipitant mix 1), H5 (buffer system 2, precipitant mix 1) and H8 (buffer system 2, precipitant mix 4 (25 % v/v MPD, 25 % w/v PEG 1000, 25 % w/v 3350)). The best hit was F12 (0.12 M Monosaccharide additives, 0.1 M Buffer system 3, 50 % v/v precipitant mix 4), which produced a rectangular structure in the 3:1 drop ratio well. F12 saw further edge refinement in the following two weeks, into a clear a rectangular formation (Figure 8-5A, Appendix B). JBS Deep Purple dye was pipetted into the edge of the droplets for Morpheus® conditions A9, B9 and C9. The dye struggled to migrate through the viscous droplet, and the aggregate formed in Morpheus A9 300:100 nL disintegrated when disturbed with the pipette. The B9/C9 aggregates disintegrated upon exposure to ambient air, and the A9 300:300 nL droplet was largely undisturbed.

3.4.2 MIDAS™ crystallisation trials

AvaL1 was also screened using the MIDAS™ screening kit at 300:100 and 300:300 nL drop ratios of crystallant:protein on the Mosquito® LCP. Initial crystallisation trials indicated that only condition D12 (20 % v/v Jeffamine® SD-2001, 15 % v/v 1-propanol) produced crystallisation of AvaL1, with 3D microcrystals clearly seen with high nucleation (Figure 8-4, Appendix B). The crystals were present in both drop volume ratios and appear larger with a higher relative ratio of crystallant. Later re-evaluation showed several new observations, with the appearance of three more aggregates from the MIDAS™ screen (Figure 8-5B, Appendix B). When the MIDAS™ D12 well was exposed to ambient air, the crystals disintegrated over the course of approximately 1 hr. However, two weeks after being sealed and allowed to re-equilibrate some crystals reformed.

The MIDAS™ D12 condition was scaled up on 24-well plates using the hanging drop method, with concentrations of 11 mg mL^{-1} for AvaL1 and 16 mg mL^{-1} for AvaL2, using a 0.2 mL reservoir liquor (Figure 3-13). The reservoir crystallant and protein solution where mixed in ratios of 1:1 µL, 1:2 µL and 2:1 µL crystallant:protein. Crystal formations where observed within days of setup. Compared to the sitting drop screen the amount of nucleation was similar at a 1:1 drop ratio, but smaller crystals were formed. When the ratio of crystallant to protein was decreased the size of the crystals increase, but with a decrease in nucleation. When the crystallant:protein droplet ratio was increased to 2:1 µL no crystals were formed, and a dark precipitate was observed. For AvaL2 only clear drops were observed at all ratios; after several weeks phase separation containing denatured protein only at a drop ratio of 1:2 was observed.

An optimising grid was setup for the crystallisation of AvaL1 in MIDAS™ D12 at 15.8 mg mL^{-1} across 8 % v/v to 24 % v/v Jeffamine® SD-2001 and 10 % v/v to 17.5 % v/v 1-propanol using a 1:1 µL droplet ratio (Figure 8-6, Appendix B). Each component was filtered using microsart 0.22 µm filters and mixed to appropriate % v/v in filtered dH_2O. The increase in protein concentration did not affect either nucleation or the size of the crystals. At 15 % v/v 1-propanol, both an increase and decrease of 4 % v/v of Jeffamine® SD-2001 from 20 % v/v produced larger crystals, although denatured protein is present at 12 % v/v, 16 % v/v and 24 % v/v Jeffamine® SD-2001. At 8 % v/v and 28 % v/v Jeffamine® SD-2001 no crystals were formed, and only denatured protein aggregates were observed. Conditions resulting in lower nucleation and larger crystal sizes are observed when only 1-propanol % v/v is lowered; 10 % v/v and 12.5 % v/v 1-propanol droplets produced similar sized crystals, but 10 % v/v had lower nucleation. When the droplets were checked 1 week later (Figure 3-14), discoloration

of the crystals and the droplet was noted in the 10 % v/v 1-propanol solution, but none in the 12.5 % v/v droplet, possibly indicating greater protein crystal stability.

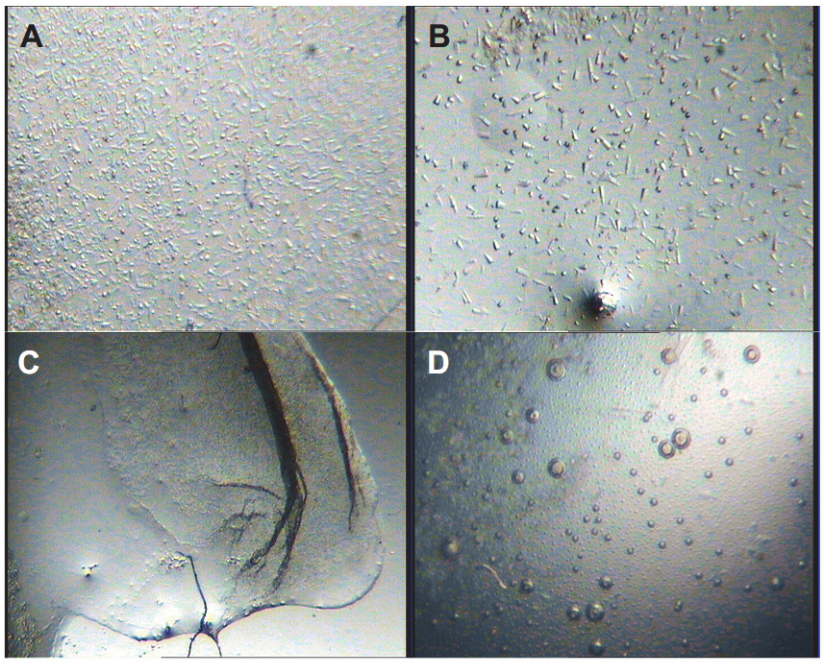


Figure 3-13 – Upscaled images from crystallisation trials of AvaL1 and AvaL2 using the stock MIDAS™ D12 condition (20 % v/v Jeffamine® SD-2001, 15 % v/v 1-Propanol) on a 24-well plate with 0.2 mL of reservoir solution. Drop ratios are indicated for crystallant:protein solution in µL. (**A**) 11 mg mL^{-1} AvaL1 at 1:1 drop ratio. (**B**) 11 mg mL^{-1} AvaL1 at 1:2 µL drop ratio. (**C**) 11 mg mL^{-1} AvaL1 at 2:1 drop ratio. (**D**) 16 mg mL^{-1} AvaL2 at 1:2 drop ratio. All images except **C** are to scale with one another. AvaL1 images collected on 12th October 2016, AvaL1 image on 1st November 2016.

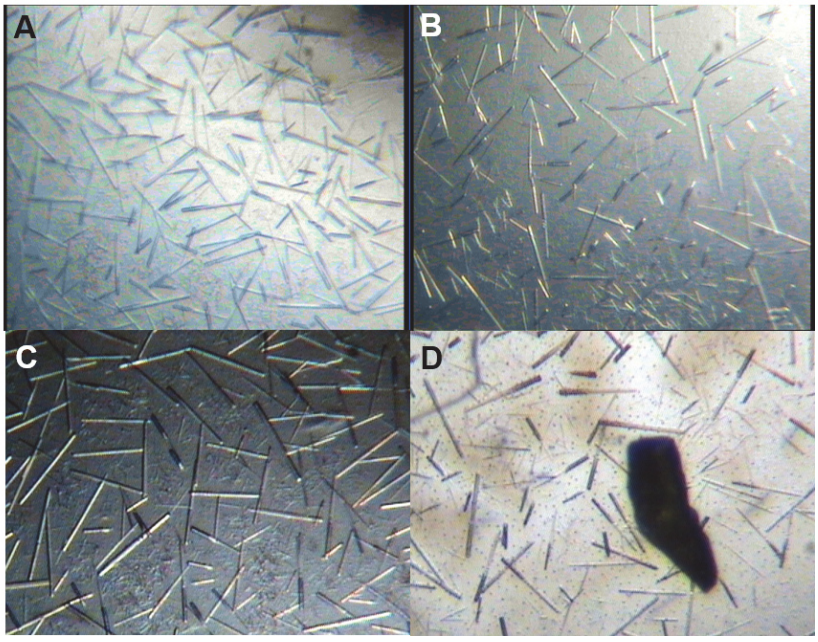


Figure 3-14 – Upscaled images of "best" AvaL1 crystal formations following optimisation of the MIDAS™ D12 condition. (**A**) 15.8 mg mL^{-1} AvaL1 in 20 % v/v Jeffamine® SD-2001, 12.5% v/v 1-Propanol and (**B**) 15.8 mg mL^{-1} AvaL1 in 20 % v/v Jeffamine® SD-2001, 12.5 % v/v 1-Propanol, after 8 days. The optimisation grid using a 24-well plate is shown in Figure 8-6, Appendix B. (**C**) and (**D**) show the same respective droplets after 15 days.

3.5 Discussion & conclusions

3.5.1 Protein preparation

All four AHFCA/pseudo-AHFCA ArpA receptor homologues from *S. avermitilis* and *S. sclerotialus* genes were successfully cloned into pET151 expression vectors. Of the four genes expressed in *E. coli* BL21(DE3), both AvaL1 and AvaL2 recombinant protein were soluble after cell lysis when expressed at 37 °C, while SclM1 was only soluble when expressed at 15 °C. SclM4$_{CO-3}$ was not soluble in the conditions tested, although at this point it is not known if the nine-amino acid C-terminal extension of this construct contributed to the observed insolubility. This means that to date AvaL2 is the only MmyR-like protein for which the biological function can be investigated *in vitro*. The ability of both AvaL1 and AvaL2 to solubilise at 37 °C would appear to indicate either greater structural stability in aqueous solution or that the ability of the proteins to fold correctly in the *E. coli* cytoplasm is greater compared to the other orthologues. Based on this both AvaL1 and AvaL2 may be more strongly suited for biotechnology applications than the other repressors.

Analysis of theoretical parameters based on the primary sequence of the proteins, including pI, amino acid composition, size, calculated stability, and calculated overall hydrophobicity using ProtParam (not shown) do not indicate any correlation with solubility, so the underlying cause for the improved stability of AvaL1 and AvaL2 currently remains unexplained. As a hypothesis, the relative insolubility of the other proteins could be due to inadequate folding of tertiary protein structure, rather than intrinsic insolubility in solution. This could be tested by determining of these proteins are functional *in vivo* in *E. coli* at high temperatures. If they are, then perhaps the change in environment upon lysis of the cell membrane may a contributing factor for insolubility due to denaturing the protein. The key to characterising solubility lies in their structure, most likely in the composition of hydrophilic and hydrophobic amino acids arranged on the solvent-accessible surface or between the interfaces of each helices; if the proteins are misfolding, it's not yet possible to say which part of the proteins are not folding correctly.

3.5.2 Assays for biological activity

AvaL1 and SclM1 bound to all the bioinformatically predicted sites with estimated similar affinities in the low hundred nM IC_{50}/K_D. The presence of the N-terminal fusion tag does not appear to interfere with the DNA-binding activity of the repressors, despite its proximity to the DBD. These results indicate that regulation of the *S. avermitilis* azoxy compound gene cluster is analogous with MmfR regulation of methylenomycin biosynthesis in *S. coelicolor* A3(2), whereby AvaL1 directly controls the expression of the transcriptional activator *saverm_2301*. SclM1 directly controls scleric acid biosynthesis by repressing the expression of all biosynthetic enzymes, analogous to the AHFCA-regulated *gbn* gene cluster in *S. venezuelae*. AvaL1 was also able to recognise *S. sclerotialus* AREs, mimicking previously observed *in vitro* binding of SgnR to $mmfR/mmfLHP_{ARE}$, suggesting that the 4 MmfR homologues share a common mechanism of DNA recognition, indicated by homology of their DBD. AvaL1 also bound two consensus sequences, one of which, *SAV_Pal_V1,* was a full palindrome (5'-AAAATACCTTCG-CGAAGGTATTTT-3'), which would be a strong candidate sequence for crystallisation of the AvaL1:DNA complex.

For the first time *in vitro*, it has been observed that one of the pseudo-AHFCA receptors, AvaL2, does not bind specifically to the same ARE sites as its partnered AHFCA receptor as previously thought, which is partially corroborated by Dr Style's *in vivo* work on MmyR. The other uncharacterised pseudo-AHFCA receptors, GbnR and SclM4, are assumed to have similar binding characteristics compared with their partner AHFCA receptors.

AHFCA6 releasing AvaL1 from $saverm\_2301_{ARE}$ indicates that despite having the least conserved LBP of all identified MmfR orthologues to date, AvaL1 is a functional AHFCA receptor. It is still currently unconfirmed if SclM1 is an AHFCA receptor, but given the greater homology between it and MmfR, this is likely to be the case.

3.5.3 Crystallography

Although crystallographic diffraction data for these recombinant proteins were not collected, crystallant conditions have been identified which allows follow-up crystallography of the AvaL1 repressor. However, the crystals produced in the MIDAS™ D12 condition may not diffract well, and there is still no confirmation that the crystals observed were indeed protein crystals. Because the proteins disintegrated when the well seal was broken, it was not possible to assess their nature using the Deep Purple dye. The crystallisation conditions themselves do not contain any salt, but the sample buffer itself has a 500 mM concentration of NaCl;

therefore, it is possible, though unlikely, that the resultant crystals are salt. However, the apparent disintegration observed when the well seal was broken may indicate the formation of protein crystals, as salt crystals tend to exhibit greater stability. The facilities to perform UV cross-polarized imaging to determine the nature of the crystals were not available. Larger crystals could be produced by using streak seeding (transferring a smashed crystal seed stock into another droplet with altered crystallant conditions; with typically 60 % to 80 % of precipitant such that crystals are not spontaneously produced), but this depends on maintaining the stability of the crystal once it is removed from the well. It is possible the presence of additives may improve the stability of the crystals in future trials. Cleavage of the N-terminal tag using TEV protease could result in better packing of the protein, given that sufficient care is taken to purify the TEV protease enzyme from the cleaved recombinant protein (potentially using a non-pET151 His_6-tagged construct). Should the crystals produced from the MIDAS™ D12 condition fail to diffract in future work, further optimisation of the Morpheus® may produce viable alternatives.

Further alteration of the buffer conditions could may also provide alternative hits. For most of the crystallant conditions that didn't precipitate protein the results were either clear-droplets or weak phase separation; the absence of glycerol in the protein buffer may have prevented precipitation, which may have otherwise revealed other potential candidate crystallant conditions. As the MIDAS™ D12 condition did not include a buffer system, pH was not altered for the stock solutions, so addition of a buffer to optimise pH could be attempted. Increasing the optimsation screen from a droplet ratio of 1:1 µL crystallant:protein to 1:2 µL crystallant:protein may further increase the size and stability of the crystals, as was seen in the pre-optimisation scaleup trials. Although not tested, SgnR might be able to crystallise in MIDAS™ D12 like AvaL1. However, AvaL1 showed no precipitation in similar conditions to MmfR, so it is possible that this may not necessarily be the case.

Because MmfR and MmfR:AHFCA2 crystallised in the same conditions, it is highly likely that AvaL1 co-bound with any of its cognate ligands will also co-crystallise in the same conditions. The conditions are likely the same due to the lack of any significant structural differences of MmfR in either its ligand-free or AHFCA2-bound state.

4 Quantifying AvaL1 DNA-binding *via* surface plasmon resonance (SPR)

4.1 Aims and Strategies

While the binding of ArpA-subfamily repressors to their cognate AREs has been widely documented in the literature (discussed in Chapter 5), quantification of these intermolecular interactions is comparatively sparse. *S. griseus* ArpA was shown to bind to an oligonucleotide (here called the ArpA consensus DNA-binding sequence, $ArpA_{CS}$) encoding a 22-bp pseudo-palindromic sequence (5'-AcAtACGGGaC|GcCCCGTtTaT-3') with an apparent affinity of 110 nM to 215 nM[204], while *S. coelicolor* A3(2) CprB was shown binding to the same sequence with an affinity of 200 nM, as well to the operator for CprB (*OPB;* 5'-gcaggCgGcaC|GgtCtGttgag-3') with an affinity of 150 nM[69, 152]. Crystallography of the CprB repressor in complex with DNA shows that the CprB:DNA complex is comprised of two pairs of homodimers bound to a single site which is not apparent in electrophoretic mobility shift assays (EMSA), with isothermal titration calorimetry (ITC) experiments indicating that this association is positively cooperative in the case of binding to *OPB* ($n_H \approx 1.6$).

In Chapter 3.3 it was demonstrated using gel shift assays that the ArpA-subfamily AHFCA receptor AvaL1 from *S. avermitilis* targets three AREs within the azoxy compound gene cluster; $saverm\_2301_{ARE}$, $avaL1/avaA_{ARE}$ and $avaL2/avaBC_{ARE}$, with an apparent affinity in the low hundreds of nanomolar, like that of ArpA and CprB. However, while gel shift assays are a good qualitative method for investigating protein-DNA interactions, determination of affinity is semiquantitative. Many physical techniques exist to quantify macromolecular interactions, including isothermal titration calorimetry (ITC), fluorescence polarisation/anisotropy (FA/FP), bio-layer interferometry (BLI) and surface plasmon resonance (SPR), as well as developing technologies such as microscale thermophoresis (MST). SPR in particular is seeing increasing prevalence for its versatility; SPR can determine not just binding affinity of macromolecular interactions but also rate kinetics, thermodynamics, stoichiometry of label-free molecular interactions over wide affinity/rate ranges in real-time, and also boasts high sensitivity/accuracy with low sample concentration requirements. However, SPR is expensive and is limited by the requirement for immobilisation of at least interaction partner on the chip surface, which may require additional labelling for some molecules. It is also not strictly suited for analysis of stoichiometric-dependent binding mechanisms. However, the nature of some systems can allow inferences to be made.

All biochemical investigations into the interactions of AHFCA and pseudo-AHFCA receptors with both DNA and their inducers have thus far been either semiquantitative when assessed *in vitro* or quantitative *in vivo*, for which other factors may influence apparent binding activity. The aim of this chapter was to quantitatively assess the DNA-binding mechanisms of AvaL1 to the endogenous ARE targets characterised previously in Chapter 3 using SPR. To take advantage of the Biacore T200 recently-acquired by the Warwick Integrative Synthetic Biology (WISB) centre, protocols were designed to collect affinity, kinetic, thermodynamic, stoichiometric and cooperativity data for AHFCA and pseudo-AHFCA receptors interacting with both DNA and ligands. Here the macromolecular interactions with DNA are investigated by injection of AvaL1/AvaL2 over a streptavidin (SA)-coated chip functionalised with 5'-biotinylated hairpin dsDNA oligonucleotides encoding the *saverm_2301*$_{ARE}$, *avaL1/avaA*$_{ARE}$ and *avaL2/avaBC*$_{ARE}$ sequences. The interactions with effectors with AvaL1/AvaL2 are reported in Chapter 6.

The theoretical principles of SPR are briefly described in Chapter 2.5.3, with definitions of important terms and principles found in Table 2-9. Note also that here the term ligand does not hold the same meaning that it would otherwise generally, instead indicating the molecule used to functionalise the chip.

4.2 Optimisation of SA-chip SPR protocol

4.2.1 Pilot experiment for binding of AvaL1/AvaL2 to DNA

Before accurate quantitative data can be collected, the Biacore system must first be optimised. A Pilot experiment for the binding of AvaL1 and AvaL2 to the *S. avermitilis* azoxy compound gene cluster AREs was carried out as described in Chapter 2.4.12. 50 µg mL^{-1} DNA solutions of *saverm_2301*$_{ARE}$, *avaL1*/*avaA*$_{ARE}$ and *avaL2*/*avaBC*$_{ARE}$ were immobilised onto an SA-chip on Fc2, Fc3 and Fc4 respectively at R_L of 1,884 RU, 1,951 RU and 1,984 RU. Each oligonucleotide (including their biotin tag) has a *mW* of 20,392 Da, 20,395 Da and 20,395 Da respectively. After buffer equilibration, 10 nM, 50 nM, 100 nM, 500 nM and 1000 nM solutions of AvaL1 and AvaL2 were each injected over the flow cells for 150 s, followed by buffer injection for 4 min. As for the gel shift assays, AvaL1 bound strongly to the three immobilised AREs with mean *BL* values of 1,456 RU ± 191 RU at 1 µM AvaL1 (Figure 4-1A), with AvaL2 having a mean *BL* value of 45 RU ± 11 RU at the same concentration (~3 % binding relative to AvaL1). The multi-cycle binding sensorgrams for both AvaL1 and AvaL2 demonstrated that the generated response was dose-dependent (Figure 4-1B). The AvaL1

binding response is multi-phasic, with two distinct association and dissociation phases that contrast with what would be expected from a single phase 1:1 Langmuir binding event. The weaker binding response of AvaL2 was more akin to a 1:1 Langmuir binding event and has a slower relative dissociation from DNA than AvaL1.

Multiphasic curves can sometimes be seen for previously determined 1:1 Langmuir binding events when the analyte sticks to the immobilised ligand as result of factors such as protein aggregation or misfolding. However, there may also be a biological inference, dependant on the mechanism of the underlying analyte:ligand interaction. AvaL1 sticking can be measured by AHFCA deregulation experiments; if the AvaL1 is sticking to DNA, AvaL1, which either would not be receptive to AHFCAs due to misfolding or non-responsive due to non-canonical binding of protein to DNA, would still be bound to relatively large amounts of DNA at high concentrations of AHFCA. This is further evaluated in Chapter 6.

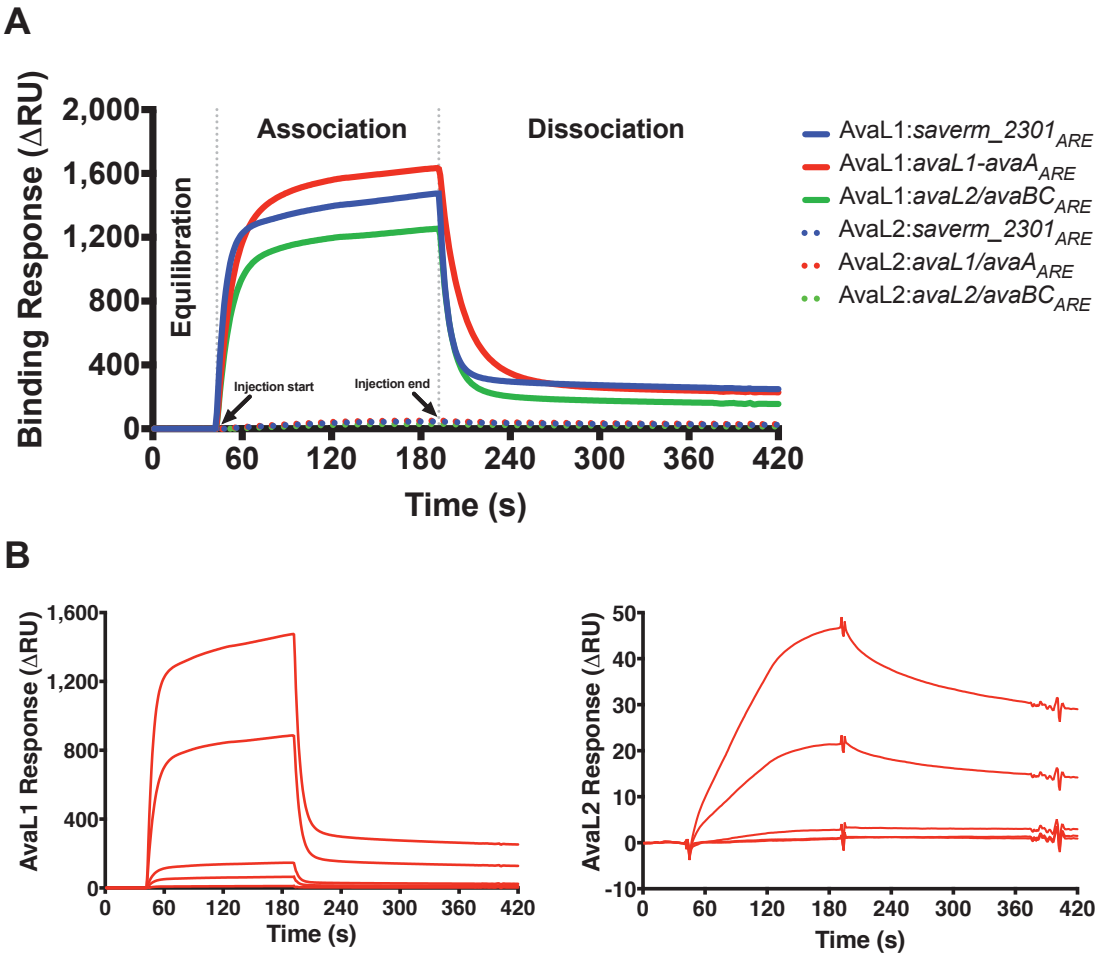


Figure 4-1 – Pilot SPR sensorgrams for the binding of AvaL1 and AvaL2 to oligonucleotides immobilised onto a streptavidin (SA)-chip. (**A**) Binding of 1 μM AvaL1 and 1 μM AvaL2 to oligonucleotides encoding *saverm_2301*$_{ARE}$, *avaL1/avaA*$_{ARE}$ and *avaL2/avaBC*$_{ARE}$ sequences, with important events annotated. (**B**) Dose-dependent binding of 10 nM, 50 nM, 100 nM, 500 nM and 1000 nM AvaL1 (left) and AvaL2 (right) to SA-immobilised *saverm_2301*$_{ARE}$. Dose-dependency is consistent across binding to all oligonucleotides (not shown). Mean R_A values are plotted with 3 replicates for each concentration, with SEM not shown due to small error values.

4.2.2 Steady state affinity modelling

When a steady state binding curve was fitted to the pilot sensorgrams, the K_D values both AvaL1 and AvaL2 binding to DNA were well above the range of assayed analyte concentration (Figure 8-7, Appendix C). The TR_{max} values for the binding of AvaL1 to DNA range from 10,264 RU to 10,810 RU (assuming a valency of 4), well above the target R_{max} of 100 RU, with a fitted R_{max} value range of 4,604 RU to 6,559 RU. The Chi^2/R_{max} and Chi^2/TR_{max} values for AvaL1 binding $saverm\_2301_{ARE}$ and $avaL1/avaA_{ARE}$ were above the significance threshold of 10 %, while for AvaL1 binding $avaL2/avaBC_{ARE}$ they were below. For AvaL2 binding the Chi^2/R_{max} values were extremely low, however with even less binding to DNA observed it is clear that the determined K_D values (with significantly large SEM values) cannot be accurate. There was insufficient analyte being injected over the chip to saturate the immobilised ligand because there is too much DNA immobilised onto the chip in this experiment for accurate model fitting. Ligand immobilisation must be optimised for not just for adequate saturation by the analyte but also because of an increase in mass transport limitations, where the rate of binding of the analyte to the ligand exceeds the rate of mass transport, resulting in rate-limitation as the protein moves to the chip surface. In addition, large amounts of immobilised ligand on the dextran surface can result in rebinding of analyte during the dissociation phase. Altogether, this indicates quantitative data cannot be accurately determined from this data. The only meaningful conclusions that can be made with these fits is that each AREs has similar affinities for each protein, and that the saturation response would appear to follow a 1:1 Langmuir Steady State model.

4.2.3 Kinetic modelling for AvaL1 binding

To characterise the observed biphasic binding of AvaL1 to DNA the models listed in Table 2-10, Chapter 2.5.3.5 were fitted to the AvaL1 sensorgrams, alongside the same models with *RI* offset constrained to 0 (Figure 4-2). The 1:1 Langmuir model fits poorly to the sensorgram and has the highest range of Chi^2/R_{max} values of all the fitted models with 164 % to 389 % across binding to all three oligonucleotides. The bivalent analyte model has a similar visual fit to the Langmuir model, however the $RI = 0$ bivalent analyte model has a better fit of 5.3 % to 11.4 % Chi^2/R_{max}, with poorer fits at lower analyte concentration cycles. The strongest fits can be seen for the heterogenous ligand and two state models, evidenced by the low residuals of the two models. Each of these models with *RI* constrained had comparatively weaker fits than unconstrained *RI* model fits. The two state model has the slightly better visual fit of the two. However, the two state model is the only model for which the Chi^2/R_{max} percentage is below the significance threshold of 10 % Chi^2/R_{max} for all three oligonucleotides.

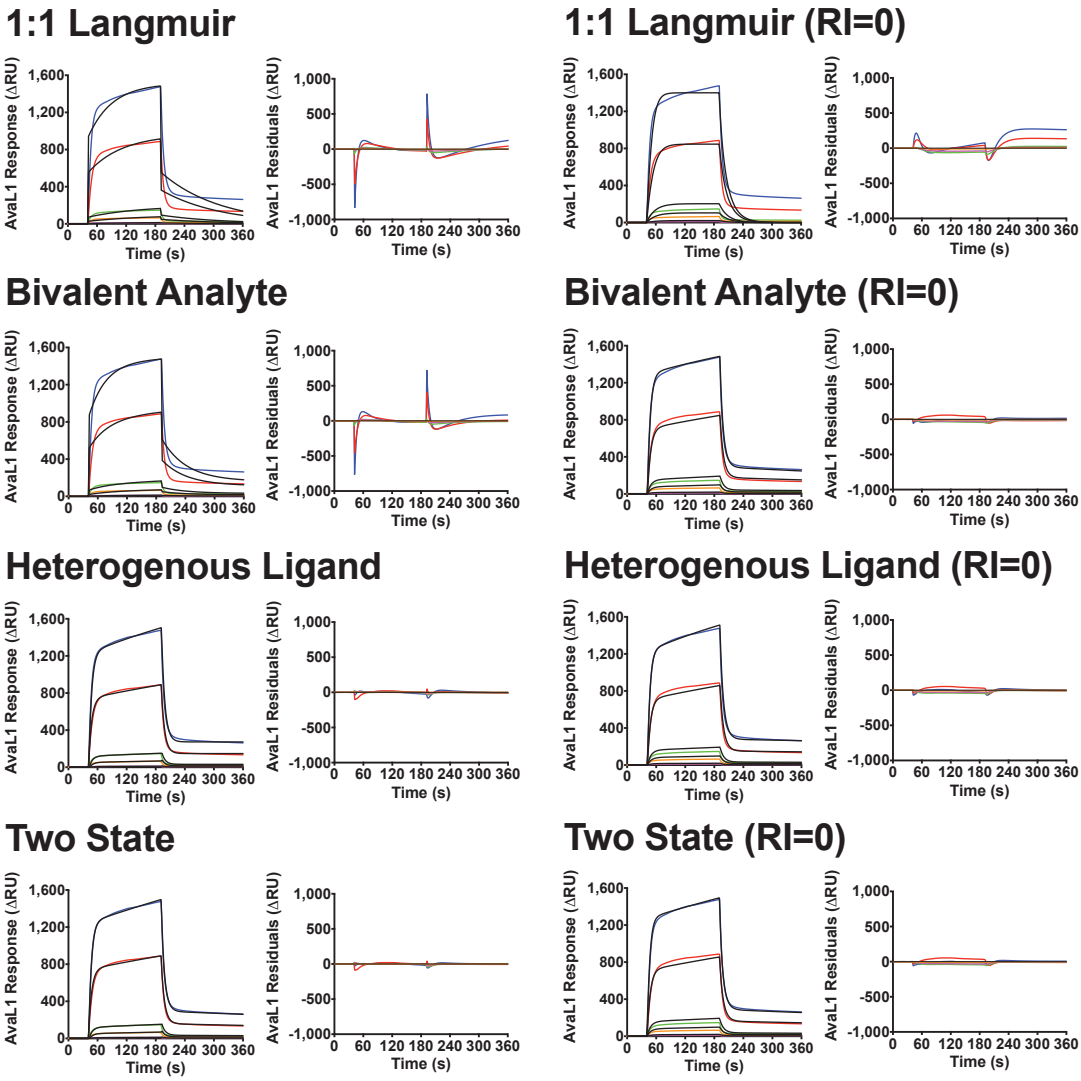


Figure 4-2 – Kinetic model fitting to SPR sensorgrams for the pilot binding of AvaL1 to SA-immobilised *saverm_2301*$_{ARE}$. AvaL1 was injected at concentrations of 10 nM, 50 nM, 100 nM, 500 nM and 1000 nM. For each model the binding sensorgrams are on the left, and model fitting residuals are on the right. Mean experimental R_A values are coloured, with the fitted model in black. Mean R_A values are plotted with 3 replicates for each concentration. SEM is not shown due to small error values. Table of model fitting parameters for binding to all three oligonucleotides can be found in Table 8-5, Appendix C.

A biological mechanism must be inferred to rationalise the fitting of the two state model. To breakdown the model components, the total response of the model is the sum of the concentration-dependent binding response component [A+L↔AL], the concentration-independent conformational shift component [AL↔(AL)], and the Bulk+Drift offset (Figure 4-3). This biphasic sensorgram is produced due to the assumed conformation of the analyte:ligand complex altering the equilibrium between bound and free forms of analyte. From analysis of TFR crystal structures a twisting conformational shift in the TFR intradimeric interface occurs upon binding of DNA, stabilising the TFR:DNA complex[64, 65]. It is proposed based on this data alone this is the same phenomenon is occurring for AvaL1. At this point, while this is easy to quantify for a conformational change with a 1:1

stoichiometry of TFR homodimer binding to DNA, it is not certain how association and dissociation can be quantified based on binding stoichiometry of two putative homodimers. As such, the two state model is currently the best representation of the observed non-1:1 Langmuir binding event. The fitting of kinetic models in subsequent SPR experiments is consistent with the findings described here, specifically regarding the poor fitting of the 1:1 Langmuir model. Therefore, differences between models is not discussed in the following sections. As was the case for the steady state affinity model fitting, the determined kinetics cannot be taken as accurate until the system is further optimised.

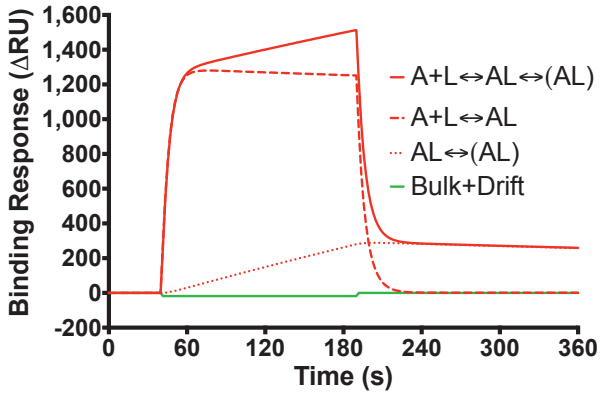


Figure 4-3 – SPR two-state kinetic model components fitted for the pilot binding 1 µM AvaL1 to SA-immobilised *saverm_2301*$_{ARE}$. A denotes analyte (AvaL1), while L denotes the immobilised oligonucleotide (*saverm_2301*$_{ARE}$).

4.2.4 Assessment of protein sample preparation methods

To determine the effects of different protein preparation methods on the ability of AvaL1 to bind DNA, SPR was performed as for the pilot sample with three sample preparations of AvaL1 from two separate protein expressions. Freshly expressed protein underwent two different buffer-exchange methods; buffer-exchange into SPR-SA Binding Buffer using centrifugal filtration (concentrated to 3.82 mg mL^{-1}; 137.7 µM), and the other buffer-exchanged *via* dialysis with a pre-injection concentration of 1.87 mg mL^{-1} (67.4 µM). Protein from a previous expression had been centrifugally concentrated and flash frozen in liquid nitrogen without glycerol and then stored at -80 °C for several weeks before the SPR experiment. Each protein preparation was diluted to a range of 0.0625 µM to 1 µM (in serial two-fold dilution steps) prior to injection. When injected over the chip (Figure 4-4A), all AvaL1 sample preparations are similarly capable of binding to the target oligonucleotides ($F(2, 42) = 0.34$, $p = 0.72$; Figure 8-8, Appendix C).

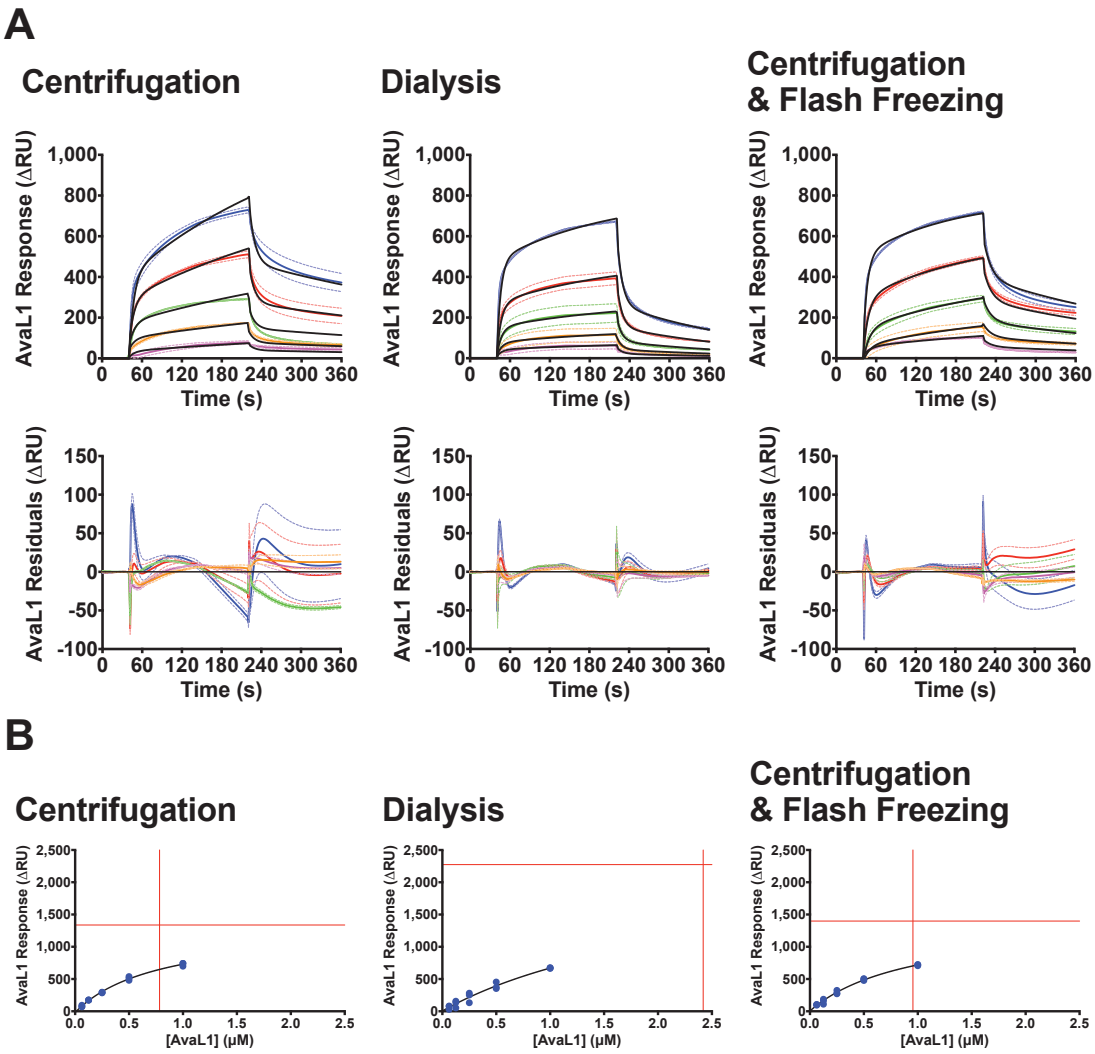


Figure 4-4 – SPR sensorgrams showing the effects of three different sample preparation methods on the binding of AvaL1 to SA-immobilised *saverm_2301*$_{ARE}$. (**A**) 0.0625 μM to 1 μM (two-fold serial dilutions) AvaL1 dose response sensorgrams for each AvaL1 preparation method were fitted with the two state kinetic model. Experimental R_A values are coloured, and the fitted model values in black. Residuals for each plot are shown below. Mean R_A values are plotted with 3 replicates for each concentration, with SEM shown by a broken line. Similar sensorgrams were observed for the binding of AvaL1 to *avaL1/avaA*$_{ARE}$ and *avaL2/avaBC*$_{ARE}$ (not shown). (**B**) Fitted saturation curves for the same data. Red vertical lines indicate K_D values, while red horizontal lines indicate R_{max} values.

Both centrifuged samples are visibly distinct from the dialysed samples, which are consistent with the original pilot experiment results. The Fitted two state model to both of the concentrated samples have larger model fitting residuals, especially in the dissociation phase. The R_A values are increased in both centrifuged samples compared to the dialysed sample, which may be indicative of protein sticking to the DNA, and dissociation of the TFR:DNA complex is also comparatively slower when centrifuged. Fitting of steady state affinity curves to each sample preparation corroborates this two state kinetics fitting (Figure 4-4B). The determined K_D values for the dialysed samples are consistent with the pilot experiment in that they are above the assayed concentration range, compared with both centrifuged samples having K_D values within the assayed concentration range. Forces applied to the protein during centrifugal filtration as it pushes against the filter membrane may result either in misfolded

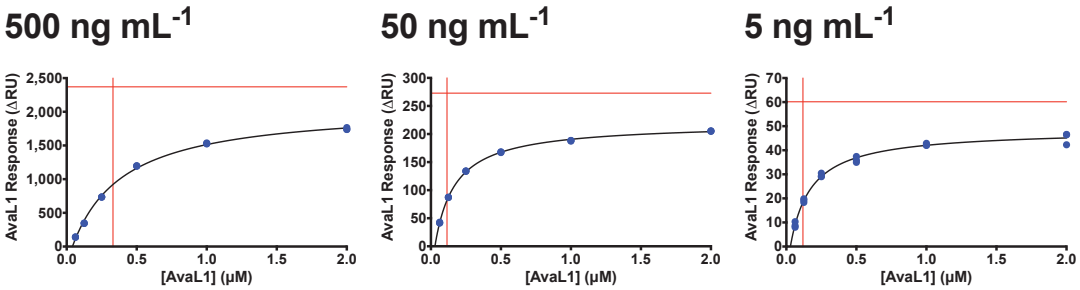
protein, soluble aggregates, or both, as all samples are briefly centrifuged in an Eppendorf tube prior to the experiment to pellet out insoluble protein. This undesirable protein may result in non-specific binding and sticking of protein to the ligand, resulting in larger R_A values and lower K_D values for both centrifuged samples. Aggregation could be tested for in future using SEC.

AvaL1 samples which were centrifugally concentrated did not show a significant change in activity when it is subsequently flash frozen, indicating a degree of freeze-thawing stability. The amount of protein binding in both samples is also a good indicator of batch consistency. Of the two centrifuged samples, the flash-frozen sample is more akin to the dialysed sample, most likely a result of batch variance rather than freeze-thawing improving AvaL1 binding. Future experiments could determine how well flash frozen dialysed AvaL1 samples can maintain their stability and activity. However, it is always best to use fresh protein buffer exchanged *via* dialysis, if time permits it. All samples used in SPR experiments from this point onward unless otherwise stated were prepared by dialysis.

4.2.5 Optimisation of DNA chip surface density

To optimise R_{max} for accurate affinity/kinetics determination, a fresh SA-chip was functionalised using 500 ng mL^{-1}, 50 ng mL^{-1} and 5 ng mL^{-1} *avaL1/avaA*$_{ARE}$ DNA solutions (diluted from 50 µg mL^{-1}) injected over each flow cell for 5 min at 10 µL min^{-1}. Binding was measured by injection of AvaL1 over the chip with a sample concentration range of 0.0625 µM to 2 µM (in two-fold serial dilutions). Upon dilution the DNA on all flow cells with lower chip surface densities are further saturated with protein compared to the pilot experiment, resulting in steady state-modelled K_D values within the range of the assayed sample concentrations (Figure 4-5). Saturation with AvaL1 still increases between 500 ng mL^{-1} and 50 ng mL^{-1} dilutions but does not change significantly between 50 ng µL^{-1} and 5 ng µL^{-1}, consistent with a decrease in the difference between calculated K_D values. The calculated K_D values for each cell surface density in order of decreasing chip surface density are (330 ± 22) nM, (115 ± 4) nM and (118 ± 17) nM, which are not significantly different than the previous EMSA-estimated affinities for AvaL1 binding specifically to *avaL1/avaA*$_{ARE}$ ($F(2, 6) = 0.06$, $p = 0.94$, Figure 8-9, Appendix C). In addition, the 50 ng µL^{-1} and 5 ng µL^{-1} immobilisations have Chi^2/R_{max} percentages of 0.8 % and 2.6 % respectively, compared to a high percentage of 41.0 % for the 500 ng µL^{-1} immobilised flow cell.

A



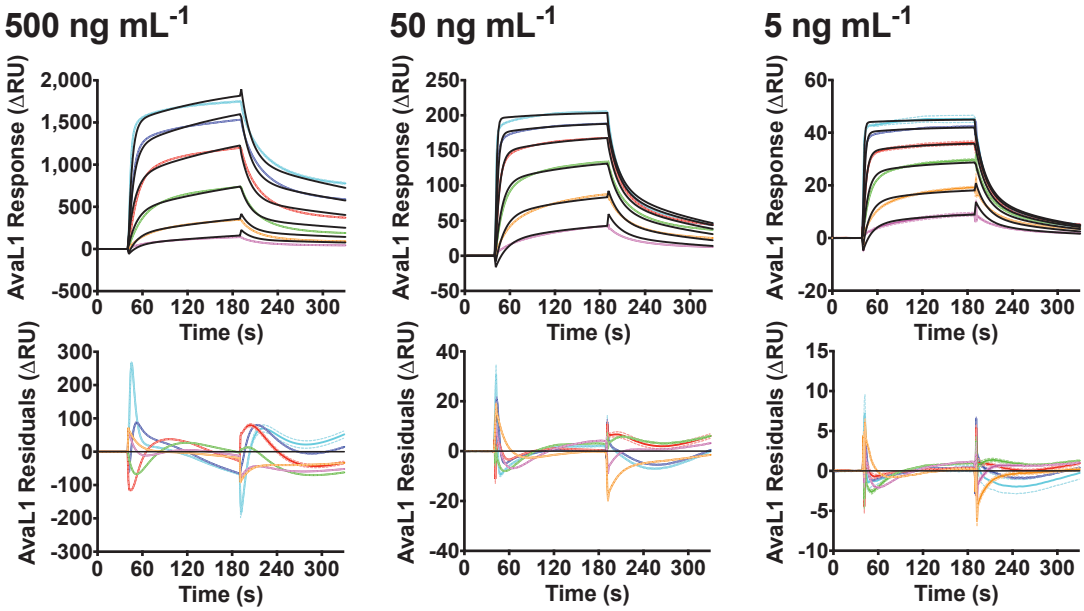
B

	500 ng mL^{-1}	50 ng mL^{-1}	5 ng mL^{-1}
K_D (M)	**(3.30 ± 0.22) x 10^{-7}**	**(1.15 ± 0.04) x 10^{-7}**	**(1.18 ± 0.17) x 10^{-7}**
Chi^2 (RU^2)	972	2.21	1.6
R_{max} (RU)	(2370.5 ± 30.0)	(272.8 ± 3.6)	(60.1 ± 2.9)
$\frac{Chi^2}{R_{max}}$ (%)	**41.0**	**0.8**	**2.6**

Figure 4-5 – SPR steady state affinity determination for AvaL1 binding to SA-immobilised *avaL1/avaA$_{ARE}$* at three DNA chip surface densities. **(A)** Saturation curves for the binding of 0.0625 μM to 2 μM (two-fold serial dilutions) AvaL1 to *avaL1/avaA$_{ARE}$* immobilised onto an SA-chip using 500 ng mL^{-1}, 50 ng mL^{-1} and 5 ng mL^{-1} oligonucleotide immobilisation solutions. Red vertical lines indicate K_D values, while red horizontal lines indicate R_{max} values. Individual R_A values are plotted for 3 replicates of each concentration. (**B**) Table of derived affinity values and fitting parameters.

In contrast to the pilot experiment, the fitted kinetic model fits are also consistent between the two state and affinity steady state models (Figure 4-6A). The kinetic model fits well to all three sensorgrams, but a bulk negative shift (difference in *RI* between the running and sample buffer) is fitted to lower AvaL1 concentrations between phases that isn't present in the binding sensorgram. The calculated K_D values for each chip surface density in order of decreasing density are 304 nM, 107 nM and 116 nM (Figure 4-6B). The kinetics values for 50 ng mL^{-1} and 5 ng mL^{-1} are of the order of 10^5 M^{-1} s^{-1} and 10^{-2} s^{-1} for k_{a1} and k_{d1} respectively, and 10^{-3} s^{-1} for both k_{a2} and k_{d2}. The k_{a1} at high chip surface density (500 ng mL^{-1}) is 10^4 M^{-1} s^{-1} but otherwise comparable with the lower chip surface densities. R_{max} values for 50 ng mL^{-1} and 5 ng mL^{-1} chip densities determined from kinetic modelling was lower than that for R_{max} derived from steady state affinity modelling, and was higher for the highest chip surface density, but is otherwise comparable. Chi^2/R_{max} values were also comparable with those derived from the steady state affinity fit, albeit generally higher except for the lowest chip surface density which had an identical Chi^2/R_{max} percentage value.

A



B

	500 ng mL^{-1}	50 ng mL^{-1}	5 ng mL^{-1}
k_{a1} (M^{-1} s^{-1})	(7.00 ± 0.08) x 10^{4}	(3.12 ± 0.06) x 10^{5}	(3.68 ± 0.03) x 10^{5}
k_{d1} (s^{-1})	(6.72 ± 0.09) x 10^{-2}	(6.75 ± 0.17) x 10^{-2}	(6.55 ± 0.06) x 10^{-2}
k_{a2} (s^{-1})	(6.02 ± 0.08) x 10^{-3}	(5.91 ± 0.11) x 10^{-3}	(5.07 ± 0.07) x 10^{-3}
k_{d2} (s^{-1})	(2.79 ± 0.07) x 10^{-3}	(5.81 ± 0.09) x 10^{-3}	(9.61 ± 0.09) x 10^{-3}
K_D (M)	**3.04 x 10^{-7}**	**1.07 x 10^{-7}**	**1.16 x 10^{-7}**
Chi^2 (RU^2)	1660.0	15.0	1.1
R_{max} (RU)	2449.4 ± 12.0	198.3 ± 0.8	42.5 ± 0.1
$\frac{Chi^2}{R_{max}}$ (%)	**67.8**	**7.6**	**2.6**

Figure 4-6 – Two state kinetic model fitting to SPR sensorgrams for the binding of AvaL1 to SA-immobilised *avaL1/avaA*$_{ARE}$ at three DNA chip surface densities. (**A**) Dose response sensorgrams for binding of 0.0625 µM to 2 µM (two-fold serial dilutions) AvaL1 to *avaL1/avaA*$_{ARE}$ immobilised onto SA-chip using 500 ng mL^{-1}, 50 ng mL^{-1} and 5 ng mL^{-1} oligonucleotide immobilisation solutions. Experimental R_A values are coloured, with fitted model values in black. Residuals for each plot are shown below. Solid lines show mean R_A values with 3 replicates for each concentration, with SEM represented by broken lines. (**B**) Table of derived kinetics and fitting parameters.

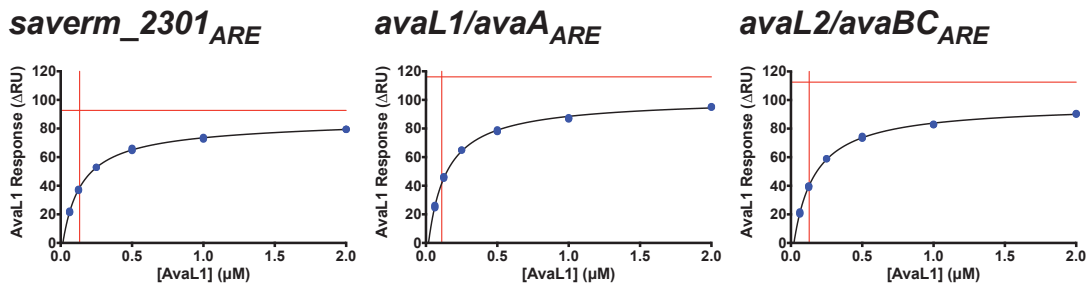
While at 5 ng mL^{-1} the R_{max} value is below the desired target value of 100 RU, the lack of any significant difference between the two lowest chip surface densities indicates that sufficient saturation has been achieved. As a result, the consistency of the affinity and kinetics for AvaL1 binding to DNA means that the chip surface density is at the optimum immobilisation level, and binding of AvaL1 to DNA should be consistent when assayed along with the other two oligonucleotides. This chip was used for further AHFCA derepressive assays described in Chapter 6.2 to assess effect of chip-surface density on AHFCA activity.

4.3 Optimised quantification of AvaL1 binding to SA-immobilised DNA

4.3.1 Steady state affinity

Each DNA oligonucleotide was immobilised onto a fresh SA-chip using 5 ng mL^{-1} solutions before binding was assayed with 0.0625 µM to 2 µM AvaL1 as for the previous experiment. As anticipated, the R_A values for AvaL1 binding were similar, with calculated R_{max} values from steady state model fitting (Figure 4-7) close to the target R_{max} of 100 RU (R_{max} range of 92.7 RU to 116.1 RU). The K_D value of AvaL1 binding to *avaL1/avaA*$_{ARE}$ is reported here as 110.6 nM ± 7.5 nM (consistent with the previous optimisation experiment), and K_D values of 132.5 nM ± 6.0 nM and 128.0 nM ± 5.7 nM for binding to *saverm_2301*$_{ARE}$ and *avaL2/avaBC*$_{ARE}$ respectively. Across all three oligonucleotides the mean K_D value is 123.7 nM ± 11.2 nM. The steady state model has a strong fit with the data, with a Chi^2/R_{max} percentage of < 1 % for binding to all three oligonucleotides. While the EMSA-derived affinity values from all three methods for semiquantitative analysis described previously (Chapter 3.3.8) were not significantly different from the SPR-derived affinities (Figure 8-10, Appendix C), EMSA-derived affinity using the AvaL1:DNA-complex bands had the least significant difference in K_D values (paired Student's $t(2)= 0.46$, $p = 0.68$), and has a similar mean K_D value of 124 nM.

A



B

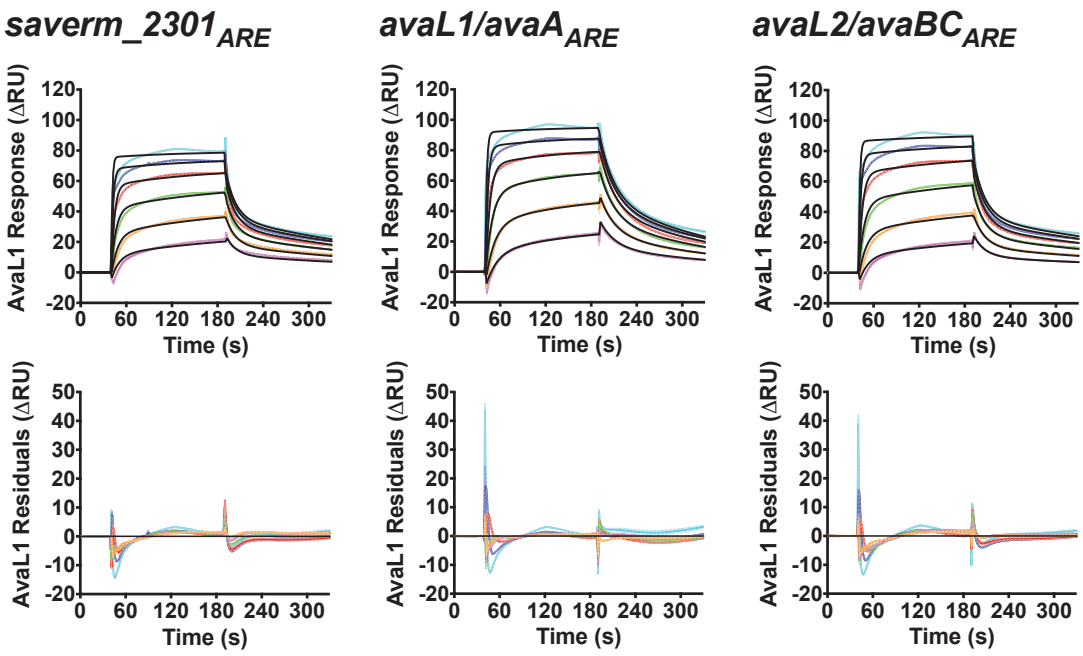
	saverm_2301$_{ARE}$	*avaL1/avaA*$_{ARE}$	*avaL2/avaBC*$_{ARE}$
K_D (nM)	132.5 ± 6.0	110.6 ± 7.5	128.0 ± 5.7
Chi^2 (RU^2)	0.39	1.12	0.53
R_{max} (RU)	92.66 ± 1.30	116.10 ± 2.80	112.50 ± 1.60
$\frac{Chi^2}{R_{max}}$ (%)	0.42	0.96	0.47

Figure 4-7 – Optimised SPR steady state affinity determination for AvaL1 binding to SA-immobilised *S. avermitilis* AREs. (A) Saturation curves for 0.0625 µM to 2 µM (two-fold serial dilutions) AvaL1 binding to *saverm_2301*$_{ARE}$, *avaL1/avaA*$_{ARE}$ and *avaL2/avaBC*$_{ARE}$ immobilised onto an SA-chip with 5 ng mL^{-1} immobilisation solution. K_D values are indicated by a red vertical line, and R_{max} values by a red horizontal line. R_A values are shown for 3 replicates of each concentration excluding 0.25 µM, for which only 2 replicates were plotted. (**B**) Table of derived affinity values and fitting parameters.

4.3.2 Two state kinetics

For the multi-cycle kinetics (Figure 4-8) there was a bulk negative shift during AvaL1 injection, which is corrected for by the two state model. 'Hump-backing' occurred during AvaL1 injection, most probably because of 'detergent effects' caused by Tween™ concentrations being below the critical micelle concentration. This effect caused less AvaL1 to be bound during the initial association phase at higher AvaL1 concentrations. Similarly, more DNA is bound at higher concentrations of AvaL1 during the dissociation phase. Because of this the model fitting had higher % Chi^2/R_{max} values (mean value of 6.05 % ± 0.83 %) compared with the steady state fitting but were still within acceptable significance.

A



B

	saverm_2301$_{ARE}$	*avaL1/avaA*$_{ARE}$	*avaL2/avaBC*$_{ARE}$
k_{a1} (M^{-1} s^{-1})	(3.51 ± 0.05) x 10^5	(2.73 ± 0.03) x 10^5	(2.86 ± 0.07) x 10^5
k_{d1} (s^{-1})	(7.03 ± 0.13) x 10^{-2}	(4.62 ± 0.07) x 10^{-2}	(6.08 ± 0.19) x 10^{-2}
k_{a2} (s^{-1})	(6.18 ± 0.12) x 10^{-3}	(5.84 ± 0.14) x 10^{-3}	(4.80 ± 0.10) x 10^{-3}
k_{d2} (s^{-1})	(3.96 ± 0.09) x 10^{-3}	(6.10 ± 0.13) x 10^{-3}	(3.31 ± 0.11) x 10^{-3}
K_D (nM)	78.16	86.57	86.68
Chi^2 (RU^2)	4.27	5.00	5.64
R_{max} (RU)	70.37 ± 0.37	95.90 ± 0.34	82.20 ± 0.46
$\frac{Chi^2}{R_{max}}$ (%)	6.07	5.21	6.86

Figure 4-8 – Optimised SPR two state kinetic determination for AvaL1 binding to *S. avermitilis* AREs. (A) Sensorgrams for 0.0625 µM to 2 µM (two-fold serial dilutions) AvaL1 binding to *saverm_2301*$_{ARE}$, *avaL1/avaA*$_{ARE}$ and *avaL2/avaBC*$_{ARE}$ immobilised onto an SA-chip at 5 ng mL^{-1} and fitted with a two state kinetic model. Experimental data is coloured, with the fitted model in black. Residuals are shown below. Solid lines indicate mean R_A values with 3 replicates of each concentration (excluding 0.25 µM for which only 2 replicates were plotted), with SEM represented by a broken line. (**B**) Table of derived kinetics and fitting parameters.

Across AvaL1 binding to the three AREs the mean kinetic values were consistent; k_{a1} = (3.03 ± 0.09) x 10^{5} M^{-1} s^{-1}, k_{d1} = (5.91 ± 0.24) x 10^{-2} s^{-1} for the forward and reverse complex formations, and k_{a2} = (5.61 ± 0.21) x 10^{-3} s^{-1}, k_{d2} = (4.46 ± 0.19) x 10^{-3} s^{-1} for the forward and reverse conformational shifts. The K_D values derived from these kinetics for AvaL1 binding to *saverm_2301*$_{ARE}$, *avaL1/avaA*$_{ARE}$ and *avaL2/avaBC*$_{ARE}$ were 78.2 nM, 86.1 nM and 86.7 nM respectively, with affinity significantly lower than for those derived from steady state affinity fitting (paired Student's $t(2) = 4.64$, $p = 0.04$; Figure 8-10, Appendix C).

4.3.3 Stoichiometry

The R_L values of *saverm_2301*$_{ARE}$, *avaL1/avaA*$_{ARE}$ and *avaL2/avaBC*$_{ARE}$ immobilised onto the SA-chip were 19.8 RU, 18.3 RU and 20.6 RU respectively. Knowing both the mW of AvaL1/DNA and the R_{max} (as determined from steady state affinity) the calculated mean stoichiometry of AvaL1 binding to all three oligonucleotides at saturation (R_{max}) were 4.04 ± 0.62 monomers per oligonucleotide (Figure 4-9). The initial hypothesis stated that AvaL1 binding should have an equivalent binding stoichiometry across all three AREs as an inherent component of the ArpA-subfamily binding mechanism based on the length of the bound sequences (Chapter 1.4.2) as well as the crystal structures of CprB in complex with DNA[69, 152]. Therefore AvaL1, like CprB, binds to DNA as homodimeric dimer of dimers.

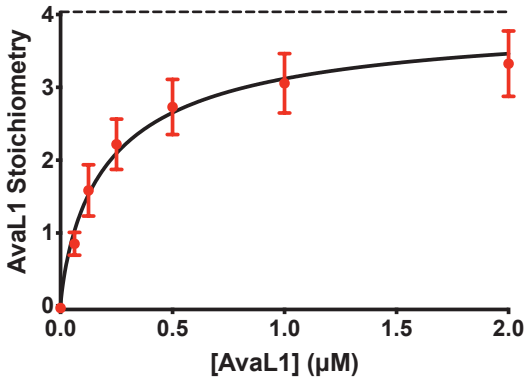


Figure 4-9 – Stoichiometry of AvaL1 monomers binding to SA-immobilised *S. avermitilis* AREs. Stoichiometry was determined using the data from Figure 4-7 and the equation from Chapter 2.5.3.6. AvaL1 stoichiometry at R_{max} (4.04 ± 0.62) is indicated by a horizontal broken line.

4.3.4 Cooperativity

n_H was determined for AvaL1 binding to all three oligonucleotides by plotting log [AvaL1] against log $\left(\frac{R_A}{R_{max}-R_A}\right)$ as described by Majka and Speck[205] (Figure 8-11, Appendix C). Using linear regression, negative cooperativity was observed for AvaL1 binding to all three AREs, with a mean n_H value of 0.82 ± 0.12. A comparison of R_A values against modelled R_{eq} values

(Figure 4-10A; with R_{eq} values calculated using the steady state fitted equation from Chapter 2.5.3.3) indicates that the Hill plot is curvilinear, with tangent slopes decreasing as the free concentration of AvaL1 increases. At AvaL1 concentrations close to the K_D value ($\log\left(\frac{R_A}{R_{max}-R_A}\right) = 0$) the tangental $n_H \approx 1$ (Figure 4-10B), indicating independent AvaL1 binding at 50 % saturation of DNA. Below the K_D concentration AvaL1 binding is positively cooperative, and above the K_D concentration AvaL1 binding is negatively cooperative. n_H values determined from the pilot and optimisation experiments show a significant correlation between the calculated n_H values and DNA chip surface density (Figure 4-10C and Figure 4-10D). This nonlinear cooperativity indicates that the cooperativity of AvaL1 binding to DNA changes as a function of the saturation of DNA by AvaL1, with cooperativity decreasing as the DNA become increasingly saturated with AvaL1.

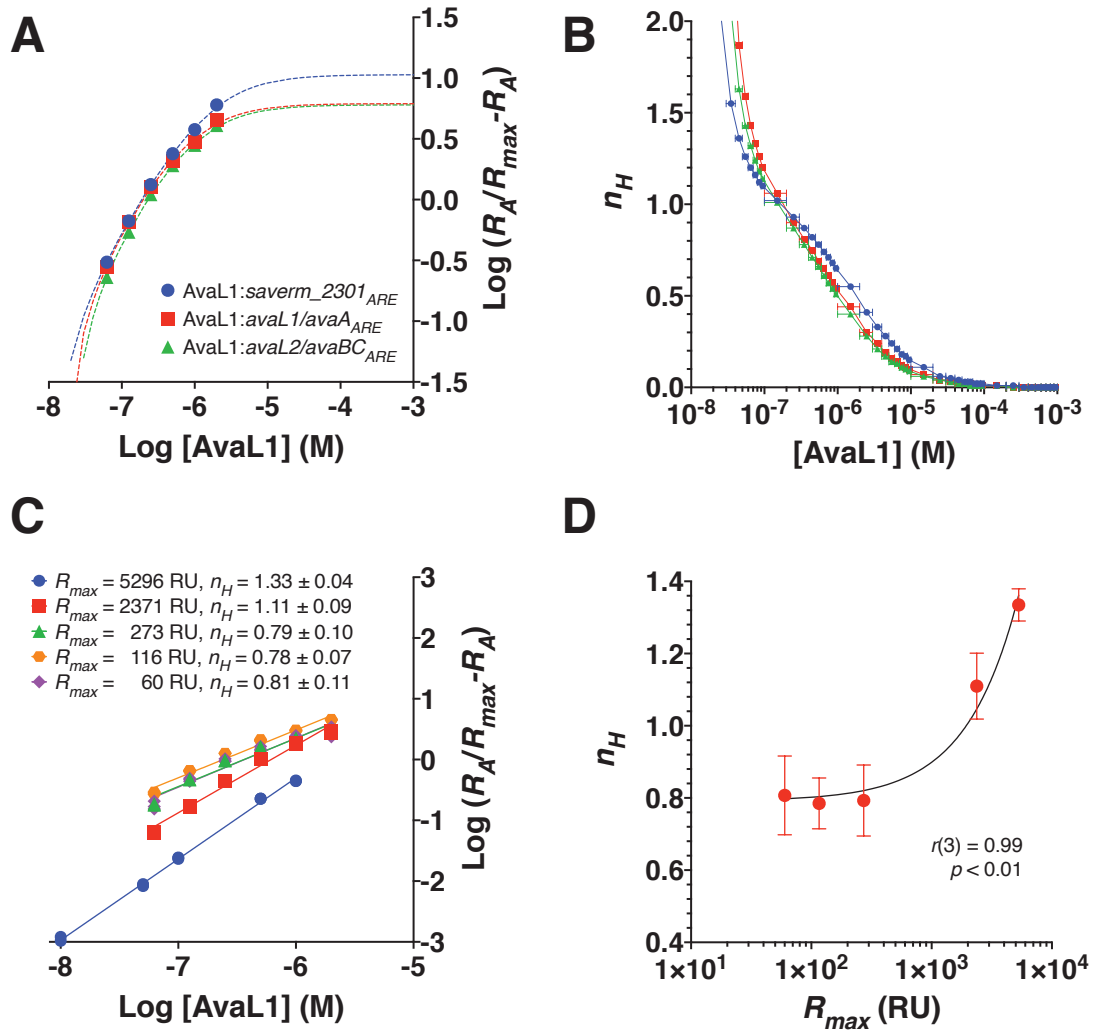


Figure 4-10 – Nonlinear cooperativity analysis for AvaL1 binding to SA-immobilised *S. avermitilis* AREs. (**A**) Hill plot using data from Figure 8-11, Appendix C with experimental values indicated by symbols (R_A) and modelled values (R_{eq}) indicated by broken lines. (**B**) Hill coefficient (n_H) plotted as a function of free AvaL1 concentration in solution, measured as the tangent slope between concentrations indicated by the x-axis error bar. (**C**) Hill plots for AvaL1 binding to *avaL1/avaA*$_{ARE}$ at different DNA chip surface densities (where R_{max} is proportional to surface density) using data from the pilot, optimisation and final experiments. (**D**) n_H values determined from **C** plotted against R_{max} and fitted with a linear regression.

4.3.5 Thermodynamics

To investigate the thermal stability of AvaL1 binding to DNA, 0.25 µM to 2 µM (two-fold serial dilutions) AvaL1 was injected over the flow cell immobilised with *saverm_2301*$_{ARE}$ from the previous experiment over a temperature range of 7 °C to 40 °C. Because of difficulties expressing the AvaL1 protein at this point, the only protein available were the samples previously centrifuged and flash-frozen. It was expected that similar issues would occur as described in Chapter 4.2.4 prior. The binding sensorgrams of AvaL1 and the range of affinities calculated from both steady state affinity and two state kinetics were consistent with these previous experiments using centrifuged samples (not shown). For steady state affinity the K_D value range was 1.5 µM to 14.5 µM, compared with slightly better fairing 0.57 µM to 1.13 µM from two state kinetics. Attempting to reproduce the saturation curves in GraphPad Prism 8 produced high parameter variability resulting in no unique fit.

Because the protein bound to DNA non-optimally, interpreting absolute values derived from the data is difficult, but useful information may be gleaned from trends in the data. There was a decrease in AvaL1 binding affinity as temperature increases (Figure 4-11). Between 7 °C and 32 °C there was a strong linear correlation between K_D (as calculated from steady state) and temperature ($r(3) = 0.92$, $p = 0.02$). However, there was a drop in the K_D value between 32 °C and 40 °C (reducing correlation to nonsignificance; $r(4) = 0.72$, $p = 0.11$). For K_D values derived from two state kinetics the same relationship occurs (for 7 to 40 °C ($r(4) = 0.60$, $p = 0.20$). This increase in affinity between 32 °C and 40 °C was reflected by a sharp decrease in k_{d1}. It is unclear whether this was because of denaturation of AvaL1 at high temperatures causing the protein to stick to the DNA or chip surface, or a measurement/model fitting error resulting from the use of centrifuged protein samples.

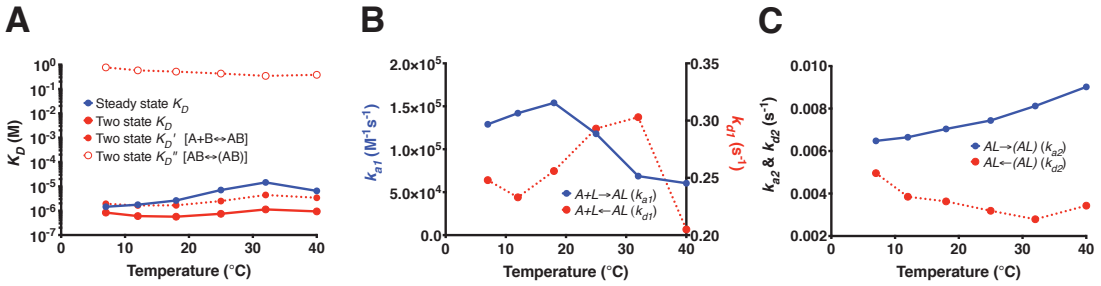


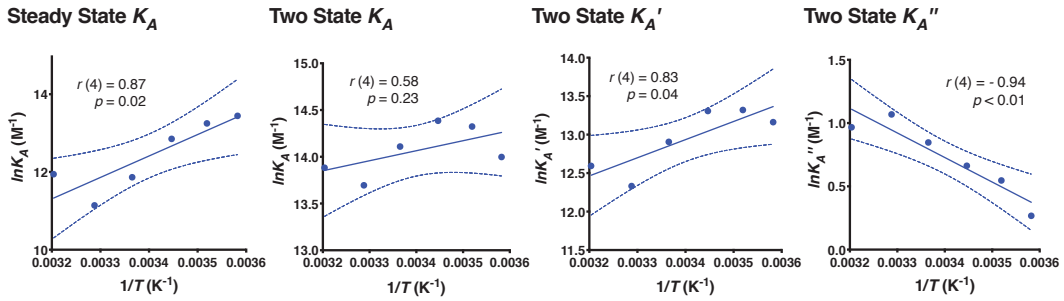
Figure 4-11 – Thermodynamic dependency of affinity and kinetics for AvaL1 binding to SA-immobilised *saverm_2301*$_{ARE}$. Binding was measured for 0.25 µM to 2 µM AvaL1 (two-fold serial dilution) to the functionalised SA-chip used in the previously experiment. (**A**) K_D and pseudo-K_D (for each step in the two state kinetic model) values plotted against temperature. (**B**) Two state rate constants for the binding reaction plotted against temperature. (**C**) Two state rate constants for the conformational shift plotted against temperature.

When the van't Hoff equation was plotted (Figure 4-12A) a positive slope was observed for both the steady state and two state models; therefore, the net reaction was exothermic (-Δ*H*). The van't Hoff plot for the binding step pseudo-dissociation constant K_D' (k_{a1} and k_{d1}) also had a positive slope, while the van't Hoff plot for the conformational shift step (K_D''; k_{a2} and k_{d2}) had a negative slope, indicating an endothermic reaction (+Δ*H*). The Eyring plots (Figure 4-12B) corroborated this, showing that all reaction steps had a negative $\Delta H^{\ddagger}$ value except for the forward conformation shift (k_{a2}), which had a positive $\Delta H^{\ddagger}$ value. That the sum of these process was exothermic may indicate that the energy released by AvaL1:DNA association contributes energy to drive forward the less feasible conformational shift.

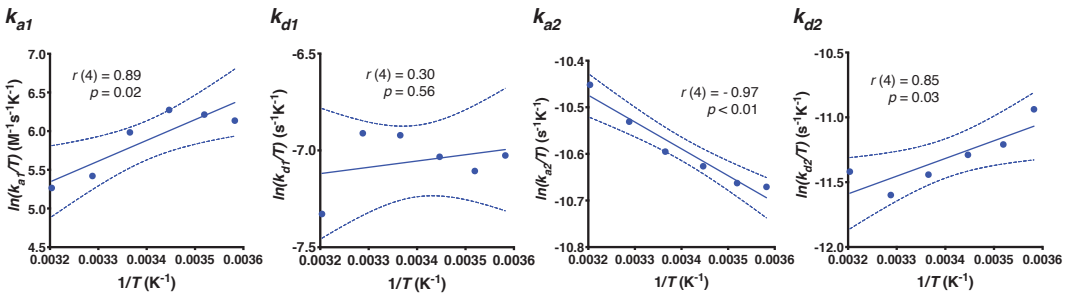
Based on the K_D values derived from steady state fitting of AvaL1 binding at 25˚C from the previous experiment, the binding of AvaL1 to *saverm_2301*$_{ARE}$, *avaL1/avaA*$_{ARE}$ and *avaL2/avaBC*$_{ARE}$ had ΔG° values of -39.25 kJ mol^{-1}, -39.70 kJ mol^{-1} and -39.34 kJ mol^{-1} respectively. Here the steady state fitting produced a ΔG° of -30 kJ mol^{-1}, with a negative ΔH° and a negative ΔS° (Figure 4-12C). As $\Delta H^{\circ} < 0$ and $\Delta S^{\circ} < 0$, the net spontaneous reaction was driven largely by enthalpy. The two state model contradicts this with a net $\Delta S^{\circ} > 0$, indicating the interaction was driven by both enthalpy and entropy; however, the van't Hoff plot for the two state model was not significant ($r(4) = 0.58$, $p = 0.23$). Based on pseudo-dissociation constants the two state model produced ΔG° values of -32 kJ mol^{-1} for the binding step and -2 kJ mol^{-1} for the conformational shift step. For the binding event $\Delta H^{\circ} < 0$ and $\Delta S^{\circ} > 0$, indicating that binding was driven by both entropy and enthalpy. For the conformational shift step both ΔH° and $\Delta S^{\circ} > 0$, indicating the less spontaneous reaction was driven by entropy.

To make the K_D values consistent with the previous experiment the R_{max} value was constrained to the TR_{max} value of 108 RU for dialysed AvaL1 binding to *saverm_2301*$_{ARE}$ and the K_D values redetermined using the steady state model (Figure 8-12A, appendix C) and two state models (not shown). the derived K_D value at 25 °C was 115 nM (compared with 131 nM from the previous experiment), while for kinetic fitting the K_D values remains relatively unchanged. The change in K_D with T was less than for the previous fitting (Figure 8-12B, appendix C), with a stronger correlation ($r(4) = 0.88$, $p = 0.02$). The previous large increase in affinity at 42 °C was not observed. The van't Hoff plot again shows that the net interaction was exothermic, with a ΔG° of -40 kJ mol^{-1}. In this model the reaction was driven by both enthalpy and entropy ($\Delta H^{\circ} < 0$, $\Delta S^{\circ} > 0$), in accordance with the previously fitted two state model (Figure 8-12C, appendix C).

A



B



C

van't Hoff	Steady state	Two state	Two state K_A'	Two state K_A''
$\Delta H°$ (kJ mol^{-1})	-46 ± 13	-9	-20 ± 7	16 ± 3
$\Delta S°$ (J K^{-1} mol^{-1})	-53 ± 44	87	41 ± 22	61 ± 10
$T\Delta S°$ (kJ mol^{-1})	-16	26	12	18
$\Delta G°$ (kJ mol^{-1})	**-30**	**-34**	**-32**	**-2**

Eyring	k_{a1}	k_{d1}	k_{a2}	k_{d2}
$\Delta H‡$ (kJ mol^{-1})	-22 ± 6	-2.7 ± 4.3	4.8 ± 0.6	-11 ± 4
$\Delta S‡$ (J K^{-1} mol^{-1})	-220 ± 20	-270 ± 15	-270 ± 2	-330 ± 12
$T\Delta S‡$ (kJ mol^{-1})	-67	-79	-80	-98
$\Delta G‡$ (kJ mol^{-1})	46	76	85	87
E_a (kJ mol^{-1})	-20 ± 6	-0.2 ± 4.3	7.3 ± 0.6	-8.9 ± 3.5

Figure 4-12 – Thermodynamic van't Hoff and Eyring plots for the binding of AvaL1 to SA-immobilised *saverm_2301*$_{ARE}$. K_A and kinetic rate constant values are derived from Figure 4-11. (**A**) van't Hoff plots for K_A and pseudo-K_A constants (K_A' for the binding step, K_A'' for the conformational shift step). (**B**) Eyring plots for two state kinetics kinetic. For **A** and **B** 95 % confidence intervals were also plotted. (**C**) Thermodynamic parameters derived from **A** and **B**. $T\Delta S°$ and $\Delta G°$ were calculated for T = 298.15 K (25 °C).

Altogether, this data suggests that even though lower temperatures produce a more stable binding response, broadly speaking there isn't a significant change in activity of the protein binding across this temperature range. This is consistent with the observed thermostability during expression of the protein; the same behaviour may not necessarily be observed for the same experiment performed on other homologues. To gain an accurate determination of binding thermodynamics the experiments should be repeated with fully functional (dialysed) protein that can either corroborate and these results. In addition, future experiments with additional temperature points could be recorded.

4.4 Discussion – context in the Literature

4.4.1 SPR-based DNA-binding of other TFRs

To determine if there was a conserved biological basis for two state model fitting, the sensorgram for AvaL1 binding to DNA was compared with literature examples of other TFRs binding to DNA which have also been assessed by SPR. SPR has been used to measure the protein:DNA interactions of *E. coli* TetR/$_{rev}$TetR[175], *Mycobacterium tuberculosis* EthR/EthR$_{G106W}$[206-208], *Campylobacter jejuni* CmeR[209], *Corynebacterium glutamicum* AmtR[210], *Salmonella enterica* serovar Typhimurium RamR[211], *M. tuberculosis* Fad35R[212], *M. tuberculosis* KstR[201], and *Mycobacterium smegmatis* AmtR[213]. Of these repressors, only the sensorgrams of TetR/$_{rev}$TetR, CmeR, Fad35R and RamR were fitted with kinetic models. TetR and $_{rev}$TetR model fitting methodology is not described in the paper, but the sensorgrams of both TetR and $_{rev}$TetR binding to *tetO* as a single homodimer appear to follow a typical 1:1 Langmuir binding sensorgram. The CmeR sensorgram was fitted with a 1:1 Langmuir model, but DNA-binding stoichiometry is not explicitly mentioned. The EthR paper by Carette X *et al* does not discuss model fitting, but the binding of EthR to the *ethA* operator (which contains two operator sites, resulting in total stoichiometry of eight monomers binding as a pair of homodimers per site) produces sensorgrams with biphasic association and dissociation phases like those produced by AvaL1. The binding of Fad35R to DNA was fitted with a single site model and confirmed to bind as a single homodimer to the *fad35R* promoter, despite being a homologue of *Bacillus halodurans* FadR which has been crystallised as a pair of homodimers binding to DNA[171].

Only the binding of RamR to P_{ramA}, collected on a Biacore T100, was fitted with the two state model. Gel shift assays indicate that RamR binds to P_{ramA} (5'-ATGAGTGctT|ActCACTCAT-3') as a pair of homodimers, which was disrupted when 2-bp in the P_{ramA} sequence were deleted (5'-ATGAGTGctT|Ac_ACTCAT -3'), as was found in an *S.* Typhimurium DT104 BN10055 MDR isolate. This disruption reduces RamR affinity from 65.8 nM to 191.3 nM, but more significantly appears to reduce the stoichiometry of RamR from a pair of RamR dimers to a single RamR dimer binding DNA. The sensorgram for this lower affinity interaction could be fitted with a 1:1 Langmuir model, indicating that the observed non 1:1 Langmuir binding of RamR may be dependent on the stoichiometry of homodimers binding to DNA. However, KstR binds to its own 18-bp operator as a single homodimer contradict this, as the sensorgram could not be fitted with a 1:1 Langmuir model; however, the sensorgrams are not shown in the paper, so it can't be said if the sensorgrams aren't like those shown for AvaL1. Were it the case that 1:1 Langmuir could be fitted to the sensorgram, the

two state model could be interpreted as a conformational shift-dependent bivalent ligand model. Ultimately, the two state model may not accurately describe the underlying mechanism of events that is occurring, as SPR is not capable of deconvoluting stoichiometric-dependent binding mechanisms, but it is clear that a change in analyte equilibrium is occurring, and therefore a conformational change is still the most probable explanation for this phenomenon.

4.4.2 Comparison of reported TFR affinities with DNA

Including AvaL1, the affinity of 22 TFRs to date have been assessed for DNA-binding affinity (Figure 4-13). In addition to those characterised by SPR as described in Chapter 4.4.1, other TFRs assayed for affinity include *Vibrio vulnificus* SmcR[214], *Pseudomonas putida* TtgR[215], *Dietza* sp. DQ12-45-1b AlkX[216], *Pseudomonas aeruginosa* CifR[217], *Myxococcus xanthus* AibR[170], *Bacillus subtilis* FadR[218], *Staphylococcus aureus* QacR[161], *E. coli* AcrR[219], *Neisseria gonorrhoeae* MtrR[220], *C. glutamicum* AcnR[221], *S. griseus* ArpA[204] and several repressors from *S. coelicolor* A3(2) including ActR/SCO7222[222], CprB[69, 152], AtrA[71] and XdhR[75].

There was no correlation observed between the stoichiometry of TFRs binding and the reported binding affinities ($r(44) = 0$, $p = 0.98$). The highest reported single affinity value for a TFR binding to its target operator is the *S. coelicolor* A3(2) *act* regulator ActR binding to the *actR/actA* intergenic region with K_D values of 0.1 nM, 0.3 nM and 5.8 nM at three different sites observed in gel shift assays. *M. smegmatis* AmtR binding to the promoter of *msmeg_2184* has the strongest affinity of DNA-binding measured by SPR, with a K_D value of 0.11 nM. The family representative TetR similarly binds to *tetO* with a K_D value of 0.18 nM, also determined by SPR. By contrast, the weakest reported affinity of a TFR is TtgR, measured *via* ITC and complimentary sedimentation equilibrium studies binding to $P_{ttgR/ttgA(wt)}$ with a K_D value of 7.1 μM. TtgR is the also the lowest reported binding for a TFR binding as a pair of dimers, with the highest reported affinity for a TFR confirmed binding as a pair of dimers being ActR, with MtrR binding to $P_{mtrCDE\text{-}long}$ with an affinity of 0.9 nM *via* FP/FA.

TFRs with the closest DNA-binding affinity to AvaL1 (123.7 nM ± 11.2 nM) are EthR binding to *ethA*-R_{ir} and CmeR binding to the *cmeABC* operon, with respective K_D values of 146 nM and 88 nM (both determined by SPR). As mentioned in the introduction for this chapter, the ArpA-subfamily repressors ArpA and CprB bind to their strongest targets ($ArpA_{CS}$, 5'-AcAtACGGGaC|GcCCCGTtTaT-3' and *OPB,* 5'-g**ca**gg**CgG**c**aC**|**G**gt**C**t**Gtt**g**a**g-3' respectively; nucleotides conserved with $ArpA_{CS}$ indicated in bold underline) with respective affinities of (110 to 215) nM and 150 nM as reported by EMSA and FP/FA. CprB binds to

ArpA$_{CS}$ with an affinity of 200 nM by FP/FA, while gel shift assays have shown that CprB binds to ScbR/ScbR2 AREs regulating *scbA* (5'-**aca**A**AC**C**G**c**a**t|tg**cC**G**GTT**ccg-3'), *cpkO*$_A$ (5'-**acA**g**AC**t**G**tgt|tag**C**t**GTtT**ca-3') and *cpkO*$_B$ (5'-**ACA**A**AC**C**gG**tg|tg**C**tG**GTTTG**T**-3') in *S. coelicolor* A3(2) with an apparent affinity between 0.75 µM to 1.5 µM[69]. CprB was also shown to bind to the exogenous *S. aureus* QacR 28-bp operator sequence (5'-**A**T**A**G**AC**C**G**AT**C**|**G**AT**C**G**GT**C**T**A**T**-3' as a 22-bp sequence in alignment with other CprB ARE sequences) with an apparent affinity of 1.5 µM[69]. The binding affinities reported here for AvaL1 binding to the three AREs in *S. avermitilis* are similar to the ArpA and CprB binding to their strongest AREs targets, and to date are the lowest reported affinity values for binding of an ArpA-subfamily repressor to DNA.

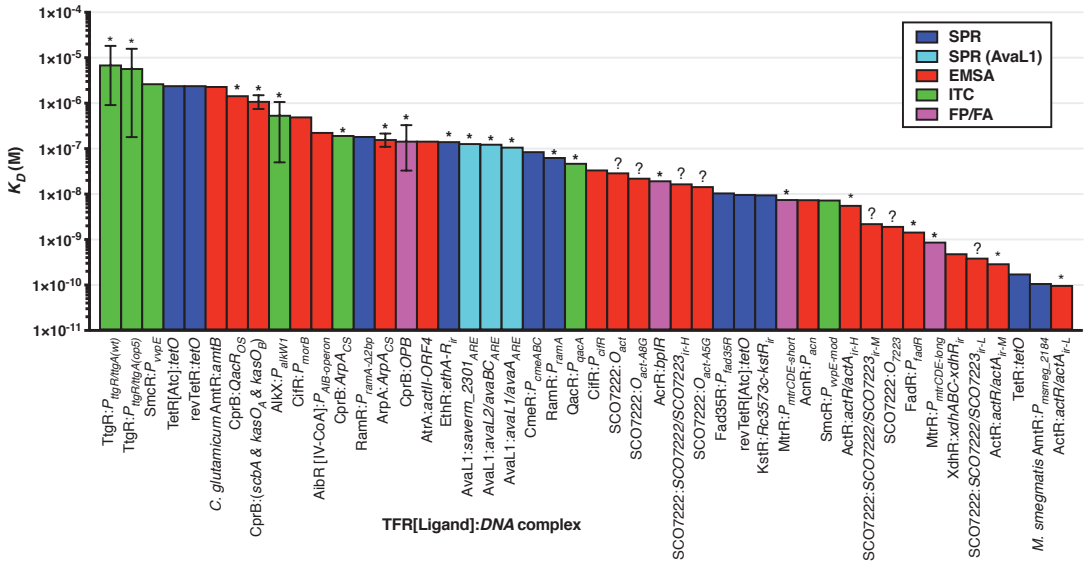


Figure 4-13 – Affinities for the binding of TFRs to target operators reported in the literature. Asterisks indicate binding to DNA as a pair of dimers, and a question mark indicates possible binding as a pair of dimers. Error bars indicate TFRs for which K_D is either estimated for overall binding or reported for individual complexes (K_D and K_D'); TtgR:$P_{ttgR/ttgA(wt)}$ = 7.1 µM (K_D = 18.2 µM, K_D' = 9.1 µM), TtgR:$P_{ttgR/ttgA(wt)}$ estimated at 5.9 µM (K_D = 15.8 µM, K_D' = 1.8 µM), CprB:(*scbA and kasO*$_A$ *and kasO*$_B$) estimated at 0.75 µM to 1.50 µM over all sequences as reported, AlkX:P_{alkW1} estimated at 436 nM (K_D = 1.06 µM, K_D' = 0.05 µM), ArpA:*ArpA*$_{CS}$ estimated at 110 nM to 215 nM, and CprB:*OPB* = 150 nM determined from FP/FA (K_D = 330 nM, K_D' = 33 nM from ITC). CifR, AibR, Fad35R, KstR and AcnR were cleaved from their purification/fusion tags before affinity was assayed (tag details in Table **8-6**, Appendix C).

ITC experiments for TtgR, AlkX and CprB report individual K_D values for each complex formation (K_D and K_D'), fitted to a sequential two site model which reports initial binding by a low-affinity complex which enables high-affinity formation of the final complex. Overall K_D values for binding are 0.78-fold the mean of the two complex K_D values (0.74-fold for TtgR:$P_{ttgR/ttgA(wt)}$ as determined by complementary analytical centrifugation and 0.83-fold for CprB as determined by FP/FA). SPR is not designed to measure formation of individual complexes (instead only measuring as a function of total binding), so it is not possible with our current data to accurately determine the affinities of sequential AvaL1:ARE complexes.

ITC experiments for AvaL1 binding would provide accurate determination of these binding sequences, assuming binding for AvaL1 is sequential.

Of the other 24 recombinant TFRs for which DNA-binding affinity was measured, 22 encode either an N-terminal (15 TFRs) or C-terminal (7 TFRs) tag (Table 8-7, Appendix C). The N-terminal tags of AibR, Fad35R, KstR and AcnR were cleaved with either thrombin or TEV protease[170, 201, 212, 221], leaving 11 TFRs with intact N-terminal tags. The N-terminal tag of AvaL1 (33 residues) is the third longest of the TFRs assayed, followed by AlkX (35 residues)[216] and the *C. glutamicum* Maltose Binding Protein (MBP)-AmtR fusion protein (393 residues), with MBP negative control experiments for the latter indicating no interaction with DNA[210]. Other TFRs expressed in pET-based vectors with uncleaved tags have a typical extension of 20-26 amino acid residues. Only in the case of AcnR is it explicitly stated that the tag was removed for the purpose of improving only protein stability, with no difference in DNA-binding activity observed between recombinant AcnR protein with intact (29 additional residues) or cleaved (2 additional residues) tags[221]. Conversely, cleavage of the N-terminal tag in MtrR decreased the stability of the protein[220]. Both ArpA and CprB were expressed in a pET26b vector which does not encode an N-terminal tag, but instead a C-terminal His_6 tag. Considering that the AvaL1 affinities are within the range of DNA-binding affinities reported for ArpA-subfamily TFRs without an N-terminal tag, and that MmfR also natively encodes a possible 25-51 amino acid N-terminal extension from the α1 helix (dependent on start codon), there is no reason to suggest that the DNA-binding ability of AvaL1 has been impacted by the presence of the His_6-V5-TEV N-terminal tag.

Some TFRs demonstrate improved binding with an increase in sequence palindromacy. Modification of the SmcR target sequence (*PvvpE*) by increasing the palindromacy (*PvvpE-mod*) improves binding from a K_D value of 2.8 μM to 76.3 nM (36-fold improvement in affinity). TtgR binding is also improved by an increase in palindromacy (*ttgR/ttgA*$_{(wt)}$ to *ttgR/ttgA*$_{(op5)}$), but to a less significant degree (approximately 1.2-fold improved activity). The converse is true for CprB, which binds to the less palindromic sequence *OPB* the more strongly as compared with $ArpA_{CS}$.

The examples of SmcR and CprB having a range of affinities demonstrates that although a TFR might have a recognition consensus that drives optimal binding to specific target, these repressors also have biologically regulated targets at sites with affinities much weaker than their optimal target sequences. Although not assessed, it is likely that the binding of ScbR and ScbR2 might have stronger affinity for sites within the *cpk* gene cluster which have closer similarity to their described recognition motifs than other 'pleiotropic targets' within the *S.*

coelicolor A3(2) genome, which may have some biological significance on regulation within the regulatory network[76]. The same may be observed for both AvaL1 and AvaL2, provided additional targets can be identified.

4.4.3 Comparing thermodynamics of ARE binding between AvaL1 and CprB

In the literature the DNA-binding thermodynamics for CprB binding to both $ArpA_{CS}$ and *OPB* sequences derived from ITC[69] corroborate some of the thermodynamics acquired for AvaL1. These ITC experiments indicate two binding modes for CprB with identical stoichiometry for each sequence (confirmed from crystal structures); single site simultaneous binding of dimers to the strongly palindromic $ArpA_{CS}$ sequence, and two site sequential binding of dimers to the less palindromic sequence *OPB*. For single site binding of CprB to $ArpA_{CS}$ ΔH = -42.3 kJ mol^{-1}, ΔS = -11.3 J deg^{-1} mol^{-1} and ΔG = -38.9 kJ mol^{-1} at room temperature. Both enthalpy and entropy are negative, as was the case for thermodynamics of AvaL1 from steady state analysis, strongly indicating that the binding of both CprB and AvaL1 to highly palindromic sequences is overall driven by enthalpy and therefore spontaneous at low temperatures.

However, formation of the CprB:*OPB* complex produces the following sets of thermodynamic parameters at room temperature:

(ΔS = -14.2 J deg^{-1} mol^{-1}, ΔH = -6.1 kJ mol^{-1}, ΔG = -37.0 kJ mol^{-1}, K_D = 330 nM)
(ΔS' = 20.5 J deg^{-1} mol^{-1}, ΔH' = 3.7 kJ mol^{-1}, ΔG' = -42.6 kJ mol^{-1}, K_D' = 33 nM)

The first interaction (for the first dimer binding) is a spontaneous event at low temperature driven by enthalpy (exothermic; ΔH and $\Delta S < 0$), while binding of the second dimer is driven by entropy (endothermic; ΔH and $\Delta S > 0$), and spontaneous at high temperature. These exothermic/endothermic sequential events mirror the individual binding and conformation steps for AvaL1, with the exception that the binding step for the AvaL1 model has a positive ΔS, thus is also driven by entropy. There is a relatively small difference in the two ΔG values for CprB:*OPB* compared with the two ΔG derived for the two state model interaction of $AvaL1{:}saverm\_2301_{ARE}$. Because the models describe two different events (for SPR stoichiometry is not a modelled variable, whereas for ITC it is), it is difficult to say whether the conformational change step described for AvaL1 binding is associated with the binding of the second AvaL1 dimer. Performing AvaL1 binding on ITC or CprB binding on SPR with the same set of protocols would determine analogy of the two modelling methods, and correlation with binding stoichiometry. It is not yet apparent whether the curvilinear Hill plot correlates with the observed binding stoichiometry of dimers to DNA.

For TtgR binding to both DNA targets both entropy and enthalpy are both positive values for each reaction step, indicating a spontaneous reaction at high temperatures[215]. It is difficult to compare TtgR thermodynamics with AvaL1/CprB as the binding affinity for its targets are so low by comparison. For AlkX binding to the *alkW1* promoter negative ΔH values are reported for both complex formations; AlkX however may be unusual in that it binds a 48-bp operator encoding two overlapping sites, which may be individuals sites for each dimer bound rather than a single dimer-dependent target[216].

4.4.4 Deconvoluting cooperativity for TFRs binding as pairs of dimers

In Chapter 4.3.4 AvaL1 binding to DNA was shown to produce a curvilinear Hill plot, indicating nonlinear cooperativity. A similar curvilinear Hill plot is also observed for the binding of EthR to DNA, showing a decrease in the tangent slopes at higher EthR concentrations[206]. However, for the tangent slope in the range of EthR concentrations close to the K_D value the reported n_H value is 1.73 (for binding to a single operator); considerably more cooperative than AvaL1. Similar curvilinear Hill plots can be observed for other multivalent ligand-receptor systems not assessed by SPR[223, 224], and linear Hill plots have also been observed in SPR experiments[225], which suggests a biological basis for curvilinear cooperativity. For the sake of comparison with other TFRs, the binding of both AvaL1 to these target AREs is interpreted as being noncooperative at concentrations around the K_D value ($n_H \approx 1$). n_H values calculated from ITC, EMSA and SPR for other TFRs binding to DNA as a pair of dimers (Table 8-7, Appendix C) indicate that binding for many other TFRs is a strongly cooperative process, with values of n_H ranging from 1.6 to 2.0. The only exception to this is ActR binding to $actR/actA_{ir}$ with an n_H value of ~1.

It is not clear whether TFRs binding as pairs of dimers to DNA do so simultaneously or sequentially. The appearances of individual dimer and pair-of-dimer complexes in some gel shift assays, with the pair-of-dimer complex becoming more prominent at higher TFR concentrations for other TFRs suggests sequential interactions with some stability of each successive dimer complex. Conversely, the appearance of single complex bands for TFRs binding as pairs of dimers may indicate simultaneous interactions where binding of a single dimer either does not occur or is not stable enough to be observed. It is probable that simultaneous and sequential modes of binding may occur for different TFRs. The binding of a TFR with $n_H > 1$ *via* simultaneous binding would suggest that binding of each dimer to DNA is interdependent, and conversely independent when $n_H \approx 1$, as was the case for ActR. A sequential model for pairs of TFR dimers binding to DNA goes as thus; the first dimer upon binding DNA undergoes a conformational rotation around the dimer interface to both stabilise

its interaction with DNA (a.k.a., the clamp-and-click mechanism), in addition to priming the DNA molecule for recognition by the second dimer. Cooperativity in this case could be measured as a function of whether this DNA priming changes the affinity of the second dimer with DNA.

The lack of cooperativity for AvaL1 binding to DNA (at [AvaL1] = K_D) and the observed models for CprB binding to *ArpA*$_{CS}$/*OPB* may not be unrelated phenomena. ITC analysis suggests that CprB binds to the *OPB* sequence through cooperative interactions (*via* a two site model), as was suggested for DNA-binding by CprB and QacR through structural analysis[69, 152, 161]. Conversely, distinct complexes for CprB binding to *ArpA*$_{CS}$ were not modelled in ITC analysis, suggesting either that each CprB dimer binds simultaneously to *ArpA*$_{CS}$ independently, or that the resulting conformational change in DNA induced by binding of the first dimer does not significantly affect the binding of the second dimer. One possible explanation for the different binding mechanisms observed for CprB DNA-binding are differences in ARE sequence palindromacy (Figure 4-14). CprB binding to *ArpA*$_{CS}$ occurs through recognition of two highly conserved (and palindromic) ARE half-dyads (5'-**AcAtACGGGaC**|**GcCCCGTtTaT**-3'), whereas *OPB* contains only one conserved ARE half-dyad (5'-gcaggCgGcaC|**GgtCtGttgag**-3'). This may suggest that one of the CprB dimers may have a poor ability to recognise the poorly conserved half-dyad, thus requiring cooperative binding with the other CprB dimer to stabilise the final CprB:*OPB* complex. For CprB binding to *ArpA*$_{CS}$ it has not been evaluated whether binding is cooperative, but for AvaL1 binding to a similarly palindromic sequence the n_H value for [AvaL1] = K_D indicates noncooperativity. This would support either simultaneous or independent sequential interactions of each dimer with each half-dyad for palindromic ARE sequences.

There are also flaws with this hypothesis; the first being that the binding of TtgR shows improved cooperativity for a more palindromically-optimised sequence (increase in n_H values from 1.63 to 1.81), rather than a decrease in cooperativity[215]. SCO7222 also binds a highly palindromic sequence (O_{7223}; 5'-CTGGAACG|c|CGTTCCAG-3') with positive cooperativity, whereas ActR binds a weakly palindromic sequence (5'-cgcGAcCA|c|CGtTCcac-3') with no cooperativity[222]. Either ArpA-subfamily repressors are an exception, which may be observed as greater structural plasticity which enables their function as pleiotropic regulators, or these different mechanisms are not dependent on sequence palindromacy but instead some other factor. More data is required to make any conclusive interpretations; no binding sites with weaker palindromacy have yet been identified for AvaL1 with which this this hypothesis may be tested. Additional experiments testing CprB binding *ArpA*$_{CS}$ and *OPB* sequences using SPR and ITC experiments for AvaL1 binding such sequences (should they exist) would be

significant in deconvoluting the binding mechanisms of these systems. It is also not clear also how this model of two pairs of dimers binding to DNA is interpreted with the observed curvilinear cooperativity of AvaL1. It would be premature to infer an accurate model for DNA-binding based on Hill plots alone, which, due to their simplicity, may not be adequate in describing binding in a system such as the one described here.

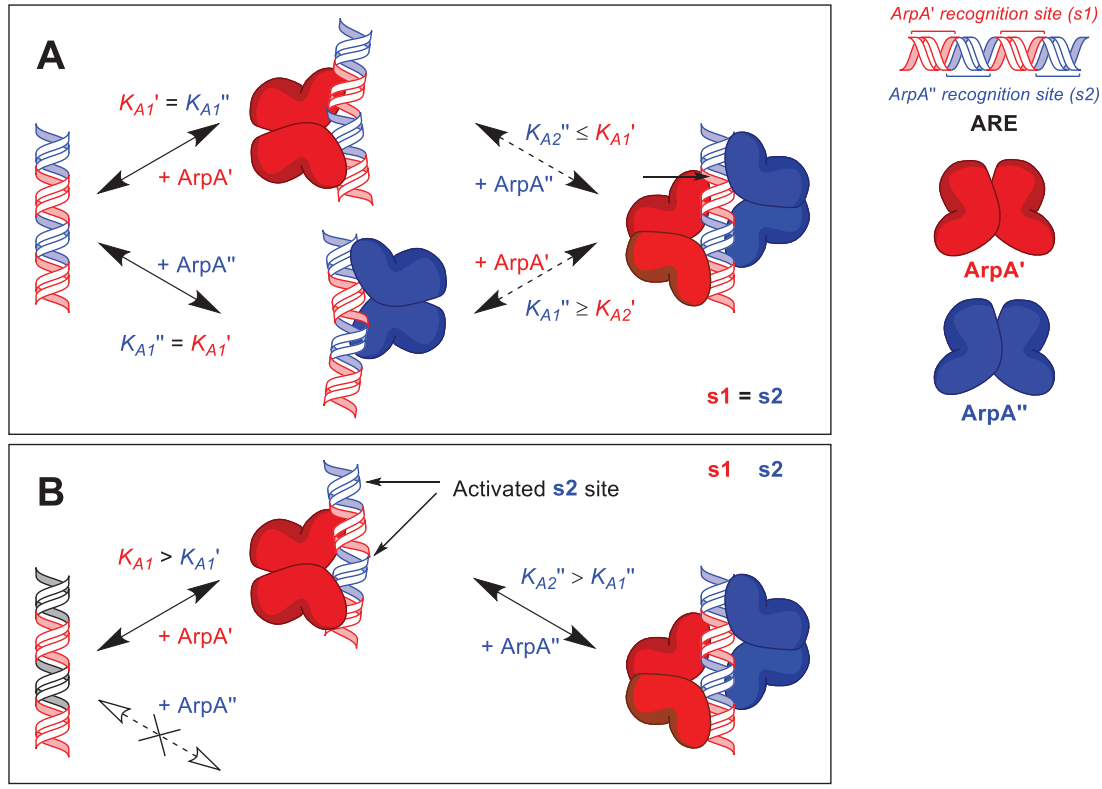


Figure 4-14 – Proposed DNA-binding mechanisms for ArpA-subfamily repressors. (A) Independent binding of dimers to a palindromic ARE, resulting in negative/lack of cooperativity ($n_H \leq 1$). This correlates with the ITC one site model observed for CprB binding to *ArpA*$_{CS}$[69] and the SPR experiments for AvaL1 binding to DNA as described in this thesis. (**B**) Necessary interdependent binding to a non-palindromic ARE, resulting in positive cooperativity ($1 < n_H < n$). This correlates with the ITC two site model observed for CprB binding to *OPB*[69]. Each ArpA-subfamily repressor dimer is represented by a prime (‘ and ‘’ for the first and second dimers respectively), and reaction steps are numbered. ArpA-subfamily repressors are hypothesised to show variable cooperativity as a function of the concentration of the repressor, with n_H interpreted as the tangent slope for repressor concentrations near the K_D value. The DNA-binding mechanisms are shown here as sequential binding events but may alternatively be simultaneous.

4.5 Conclusions

SPR corroborates the EMSA results shown in Chapter 3.3.2 showing that AvaL1 binds to three AREs upstream of *saverm_2301*, *avaL1/avaA* and *avaL2/avaBC*, and that AvaL2 does not bind these sequences. Steady state analysis shows that AvaL1 binds to its three cognate ARE targets with a mean K_D value of 123.7 nM ± 11.2 nM and binding most strongly to the ARE upstream of its own gene with an affinity of 110.6 nM ± 7.5 nM. These affinities are consistent with reported affinities of other ArpA-subfamily repressors in the literature. AvaL1 was also shown to bind DNA stably over a temperature range of 7 °C to 40 °C, and thermodynamic analysis indicate the net reaction for DNA-binding is spontaneously exothermic and driven by enthalpy ($\Delta H < 0$).

The difference in affinity between AvaL1 binding to $avaL1/avaA_{ARE}$ and $saverm\_2301_{ARE}/avaL2/avaBC_{ARE}$ is less significant than the differences in derepressed transcript levels observed *in vivo* for the repression of MmfR by either S1 nuclease protection analysis[67] or Lux-reporter assays[149]. As azoxy compound gene cluster transcripts have not been investigated *in vivo*, it may be that transcript levels of AvaL1-regulated targets are different than those regulated by MmfR in the *mmy* gene cluster. Alternatively, as optimisation of these ARE sequences in has *E. coli* shown, the promoter strength can be manipulated without significantly affecting the binding of MmfR (Dr Rodríguez García, personal communication). Therefore, observed differences in expression levels may instead arise from differences in the promoter strength resulting in different levels of RNAP binding, rather than different affinities of the AHFCA receptors for target AREs.

The binding of AvaL1 doesn't follow typical 1:1 Langmuir binding, instead showing distinct biphasic association and dissociation phases. A two state model produced the most statistically significant fit, describing a mechanism involving a binding step and a conformational shift step to form a stable AvaL1:ARE complex with altered equilibrium between AvaL1 in its bound and free forms. This two state kinetic analysis produced four rate constants; k_{a1} = (3.03 ± 0.09) x 10^5 M^{-1} s^{-1}, k_{d1} = (5.91 ± 0.24) x 10^{-2} s^{-1}, k_{a2} = (5.61 ± 0.21) x 10^{-3} s^{-1}, and k_{d2} = (4.46 ± 0.19) x 10^{-3} s^{-1} for the forward and reverse conformational shift steps (shown as the mean of binding to all three AREs). The binding step was driven by both enthalpy and entropy ($\Delta H < 0$, $\Delta S < 0$, while the second step was an endothermic reaction driven by entropy ($\Delta H > 0$, $\Delta S > 0$). This agrees with CprB thermodynamics, which also indicates the second step reaction may be synonymous with the binding of the second dimer to the target ARE.

The two state model appears to correlate with the binding of a pair of dimers to DNA observed in the literature, and stoichiometric analysis shows that AvaL1 also binds to DNA as a pair of dimers (4.04 ± 0.62 AvaL1 monomers at R_{max}). This may suggest that the conformational shift observed for binding is linked with the association of two dimers with DNA. However, the binding of these AvaL1 dimers to DNA was non-cooperative ($n_H \approx 1$) at AvaL1 concentrations close to the K_D concentration and appear to indicate a curvilinear cooperativity. From comparisons with CprB it is here proposed that the observed cooperativity of ArpA-subfamily repressors binding to DNA is dependent on the palindromacy of the target sequence, although this hypothesis requires further investigation in future experiments.

A series of protocols *via* optimisation of sample preparation and DNA immobilisation level have here been established for follow-up analysis of AvaL1 and AvaL2 using SA-chip-based SPR. It is hoped in the future that the DNA-binding of homologous AHFCA and pseudo-AHFCA receptor proteins, as well as other ArpA-subfamily repressors, will be investigated with regards to their biological function as well as use in biotechnology. These methodologies are applicable to other TFRs, including many orphan TFRs from *Streptomyces* which may be of future interest to WISB or collaborating partners. Plans were initially made to perform binding analysis with ITC to compliment the data collected by SPR, but technical issues prevented ITC experiments from being performed.

5 Characterising ARE sequence recognition and DNA-binding mechanisms of AHFCA receptors

5.1 Aims and Strategies

The Corre group has previously demonstrated how the AREs of AHFCA receptors in previously uncharacterised *Streptomyces* species can be accurately predicted using the binding sites of known repressors. The MmfR palindromic ARE consensus sequence in the *S. coelicolor* A3(2) *mmy* gene cluster was initially identified using bioinformatics and comparison with other ArpA-subfamily repressor binding sites[67]. This MmfR motif was then used to not only correctly predict *S. venezuelae* SgnR AREs in the *gbn* gene cluster[148], but has now identified *S. avermitilis* AvaL1 AREs in the azoxy compound gene cluster and *S. sclerotialus* SclM1 AREs in the *scl* gene cluster. Recent developments demonstrate that ArpA-subfamily repressors which were once thought only to be specific regulators of the gene clusters in which their genes are encoded, such as for *S. coelicolor* A3(2) *cpk* regulators ScbR and ScbR2, are now understood to have pleiotropic regulatory roles as well[76]. Since then, the ArpA-subfamily gene cluster repressors AvaR1, AvaR2 and SbbR have since also been shown to demonstrate both gene cluster-specific and pleiotropic regulatory activity[77-79].

New efforts are being made in identifying these pleiotropic targets, through bioinformatic means or otherwise. Dr Poon has previously used bioinformatic tools such as the Multiple Em for Elucidation of Motifs (MEME) Suite for prediction of ARE sequences in multiple *Streptomyces* spp., in addition to phylogenetics to characterise AREs within the context of all AREs found in *Streptomyces* AHFCA-regulated gene clusters[112]. A better understanding of these AREs should enable stronger predictive power to identify these pleiotropic sites; however there has also not yet been a comprehensive comparison of AREs bound by ArpA-subfamily repressors, especially between the AREs bound by AHFCA receptors and other repressors. In addition, little work has been done to describe how the protein-DNA interactions of ArpA-subfamily repressors relate to their binding mechanism and specificity, especially for AHFCA and pseudo-AHFCA receptors. Much of what is known about the molecular interactions between ArpA-subfamily repressors and DNA comes from the crystal structures of the *S. coelicolor* A3(2) pleiotropic regulator CprB, which is the only member of this subfamily for which structures of the repressor in complex with DNA are available[69, 152].

This bioinformatics-focused chapter is split into two parts; Chapter 5.2 reviews and characterises the AREs bound by *Streptomyces* AHFCA receptors and contextualises them in comparison with AREs bound by other ArpA-subfamily repressors using MEME Suite bioinformatic tools. The predictive methodologies used by Dr Poon is also expanded upon to predict pleiotropic targets of AHFCA/Pseudo-AHFCA receptors, while simultaneously considering the possibilities of multiple ArpA-subfamily repressor systems working in tandem with one another. It is hoped that this will provide insightful for future experiments. Chapter 5.3 is focused on structural bioinformatics, seeking to investigate *in silico* the protein-DNA interactions formed by the AHFCA receptor MmfR using the crystal structure of MmfR and CprB as references, and how these interactions relate to the observed two state mechanism observed for *S. avermitilis* AvaL1 binding to AREs in Chapter 4. All sequences used for MEME Suite analysis are available in the Electronic Supplementary Materials.

5.2 Bioinformatics analyses of MmfR-type AREs

5.2.1 MEME-Suite: tools for bioinformatic analysis

To reiterate from Chapter 1.4.2, sequence palindromacy is thought to be necessary for orientating and stabilising the binding of TFR dimers to DNA in the case for each monomer of a single dimer binding such as TetR[157], and for each dimer in the case of TFRs binding as pairs of dimers, such as CprB[69, 152]. A comparison of motif Expect (*E*) values between a nonpalindromic and palindromic motif can be used as an indicator of the strength of palindromacy of a given sequence bound by an ArpA-subfamily repressor, with a low difference indicating strong palindromacy. A nonpalindromic MEME motif is calculated by analysis of both positive/negative-sense of each sequence within a given sequence dataset, with scoring variables calculated for the sequence sense best fitting the motif enriched across all sequences. For palindromic motifs the nonpalindromic motif is first calculated, then the nucleotide frequency at complementary positions around the dyad centre are averaged (positions 1 and 24, 2 and 23, *etc*; relative positions around the dyad centre are represented as -12 and +12, -11 and +11, *etc*.). The *E* value simply put is defined as the number of hits of a motif scoring equally or greater than the motif shown that can be expected to occur by chance if all nucleotides sequences within a database of a similar size were randomised. *E* values decrease with an increasing number of sequences and increasing length of each sequence within the dataset. Thus, the *E* value can be interpreted as a measure of significance much like the *p* value for statistical tests with a standard significance cut-off of 0.05. *p* values are also given for each site found within the input sequence, defined as the probability that an equal or

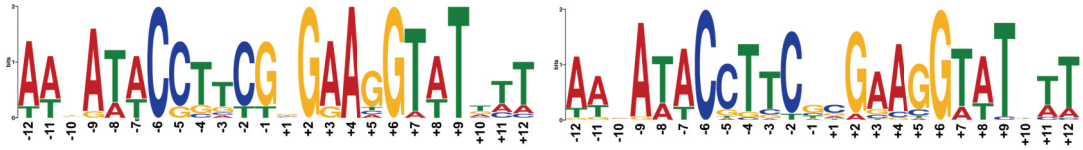
better site (based on match-scoring to the motif) would occur in a randomised sequence of the same length. Thus, the *p* value serves as a measure of statistical closeness to the given motif. Palindromic and nonpalindromic motifs can be compared against one another using the MEME Suite tool Tomtom[226], which can report useful comparative statistics including *p*/*E*/*Q* values; in this work the *p* value is used, defined in Tomtom as the probability that a random motif of the same width as the target would have an optimal alignment with a match score as good or better than the query with the target. An alternative assessment can be made by comparing the ratio of $\log E_{nonpalindromic}/\log E_{palindromic}$ values, with a ratio closer to 1 indicative of stronger match between the palindromic and nonpalindromic motifs.

5.2.2 Characterising the MmfR-type recognition motif

Using MEME, palindromic and nonpalindromic ARE motifs were generated from all sites bound by the MmfR orthologues from *S. coelicolor* A3(2), *S. avermitilis*, *S. venezuelae* and *S. sclerotialus* (Figure 5-1). Based on previously observed binding of *S. coelicolor* A3(2) AREs by both MmfR and SgnR[139, 148], and the binding of *S. sclerotialus* AREs by both AvaL1 and SclM1 observed in Chapter 3.3, it is assumed that each MmfR orthologue can recognise all ARE sequences from the four *Streptomyces* spp., thus sharing a common DNA recognition mechanism. Synthetic consensus sequences also bound by AvaL1 (Chapter 3.3.3) were not included as they themselves were derived from these generated motifs. The nonpalindromic motif has a consensus of 5'-AAnATACCTTCG|bGAAGGTATdTT-3' (b = C/G/T, d = A/G/T and n = any nucleotide) with an *E* value of 1.0^{-63}, and the palindromic motif a consensus of 5'-AAnATACCTTCG|CGAAGGTATnTT-3' with an *E* value of 5.6^{-59}. Both motifs are statistically significant, with a significant comparative Tomtom *p* value of 3.7^{-13} and log *E* value ratio of 0.92 indicating that the MmfR-type binding motif is strongly palindromic.

Although all of the AREs are palindromic to varying degrees, none are full palindromes. The most palindromic sequence is *avaL1/avaA*$_{ARE}$, with only a single nucleotide mismatch at position ±3 around the dyad centre. The least palindromic ARE is *sclM4/sclM5*$_{ARE}$, for which only half the nucleotides are palindromic. Against the palindromic motif the AREs with the lowest *p* values are *avaL1*/*avaA*$_{ARE}$, *mmyR*$_{ARE}$ and *mmfR*/*mmfLHP*$_{ARE}$, and the highest *p* values *sclM4*/*sclM5*$_{ARE}$, *avaL2*/*avaBC*$_{ARE}$ and *gbnR-4186*/*sgnH*$_{ARE}$. However, despite *avaL1/avaA*$_{ARE}$ and *sclM4/sclM5*$_{ARE}$ being the least similar to one another gel shift assays showed no significant difference in AvaL1 binding between these two AREs or any of the other assayed AREs (Chapter 3.3). Similarly, the SPR experiments for AvaL1 binding to *avaL1*/*avaA*$_{ARE}$, *avaL2*/*avaBC*$_{ARE}$ and *saverm_2301*$_{ARE}$ (Chapter 4.3.1) indicate no significant correlation between *p* values and reported affinity ($r(1) = 0.46$, $p = 0.69$).

A



Nonpalindromic motif: $E = 1.0^{-63}$ Palindromic motif: $E = 5.6^{-59}$

B

ARE	Sequence (5' to 3')	p_{np}	p_p
Nonpalindromic Motif	AAnATACCTTCG|bGAAGGTATdTT		
Palindromic Motif	AAnATACCTTCG|CGAAGGTATnTT		
FIMO Best Match	AATATACCTTCG|CGAAGGTATATT		
***S. coelicolor* A3(2)**			
mmfR/mmfLHP$_{ARE}$	AatATACCTgCg|gGaAGGTATTaT	7.07^{-12}	1.24^{-10}
mmyR$_{ARE}$	AACATACCcTCg|gGAaGGTATGTT	9.62^{-12}	1.09^{-10}
mmyB$_{ARE}$	gtaAAACCTTCc|cGAAGGTTTttt	7.55^{-10}	2.49^{-9}
S. avermitilis			
avaL1/avaA$_{ARE}$	AATATACCTgCG|CGaAGGTATATT	1.13^{-11}	9.55^{-12}
avaL2/avaBC$_{ARE}$	AgaATtCGTTCa|cGAACGtATctT	7.78^{-9}	2.94^{-8}
saverm_2301$_{ARE}$	AtAATtCCTTtg|agAAGGtATTtT	1.00^{-9}	4.77^{-9}
S. sclerotialus			
sclM1/sclM2LM3$_{ARE}$	AATATACCTatg|ggaAGGTATATT	6.22^{-10}	8.07^{-9}
sclM4/sclM5$_{ARE}$	AAagAtCcgCCt|tGGaaGtTtcTT	7.52^{-7}	9.54^{-6}
sclQ1-4P$_{ARE}$	tTgAAACCTTCG|CGAAGGTTTgAt	1.10^{-9}	8.33^{-10}
sclN-E$_{ARE}$	tgaATaCCTcCG|CGaAGGaATgat	1.32^{-9}	3.99^{-9}
S. venezuelae			
sgnR/sgnL-4184-sgnP$_{ARE}$	AATATACaTTCt|cGAAgGTATATT	5.10^{-10}	1.40^{-9}
gbnR-4186/sgnH$_{ARE}$	AAgATtCggTCt|gGAagGtATtTT	6.87^{-10}	1.45^{-8}
gbnABC$_{ARE}$	TtgAtACCTTCg|tGAAGGTtTttA	8.32^{-10}	5.20^{-9}

Figure 5-1 – 24-bp sequences and motifs recognised by four recombinant *Streptomyces* MmfR-type orthologues. (**A**) Nonpalindromic and palindromic motifs generated using all 13 AREs. (**B**) Table of all 13 confirmed ARE sequences with statistics of match to the non-palindromic and palindromic motifs. Nucleotides in uppercase indicate palindromacy around the dyad centre. All *E* and *p* values are derived from MEME analysis. p_{np} indicate *p* values for matching of each sequence to the nonpalindromic motif, and p_p the *p* values for matching to the palindromic motif. b = C/G/T, d = A/G/T, and n = any nucleotide,

As with other ArpA-subfamily repressor AREs, the distal regions are largely occupied with A/T nucleotides, with proximal positions ±6, ±5, ±2 and ±1 occupied with predominantly C/G nucleotides. The binding motif can be interpreted as blocks of A/T and C/G dinucleotides (Figure 8-13A and Figure 8-13B, Appendix D); with a dinucleotide block consensus of 5'-WWWSWS|SWSWWW-3' (W = A/T and S = C/G). This dinucleotide configuration corroborates with the frequency of observed nucleotide occurrences ($r(4) = 0.98$, $p < 0.01$), with the ±4 dinucleotide position always occupied by AT-only nucleotides. This is reflected in the GC content, for which the 24-bp MmfR binding motif has a mean G/C content of 36.8 % ± 7.0 %; approximately one-third content by G/C. In the strongly conserved TATA region described in other ArpA-subfamily repressor motifs, positions ±10 (**T**ATA…TAT**A**) in the MmfR-type motif may be occupied by any nucleotide. Similarly, the dyad centre nucleotide positions ±1 exhibits similar degeneracy, although C/G nucleotides are the most frequent nucleotides in these positions. In the nonpalindromic motif only five nucleotide positions out

of 24 are conserved across all sequences; C-6, G+2, A+4, G+6, and T+9, of which C-6 and G+6 are the only nucleotides that pair together in palindromic fashion around the dyad centre (thus the only nucleotides with full conservation in both motifs). Nonpalindromic motif nucleotides positions -12, -11, -8, -7, +4, +7, +8 and +9 are exclusively occupied by A/T nucleotides (±7 and ±8 only in palindromic motif), with positions -9 and +3 always encoding an A/G purine base (in the palindromic motif -9 is purine base, +9 is a C/T pyrimidine base, and -3 is no longer purine only), and position -2 always encoding a pyrimidine base (with +2 also a complementary purine base in the palindromic consensus).

As mentioned in Chapter 1.2.3, other putative AHFCA-regulated systems are found in other actinomycetes bacteria. Figure 5-2 shows a palindromic motif built with an input total of 42 sequences, including the thirteen endogenous AREs and the two consensus AREs characterised by gel shift assays previously. Because the MmfR binding motif was initially used to identify these sequences, it is not surprising that many of the motifs identified in these species are similar in palindromacy and composition. However, the putative *S. hygroscopicus* subsp. *jinggangensis* 5008 AREs have high C/G nucleotide frequencies in the ±8 and ±7 positions normally occupied by AT nucleotides, and the ±6 position is no longer fully conserved by C/G nucleotides. When parsed through MEME two palindromic motifs were produced; one without the *S. hygroscopicus* subsp. *jinggangensis* AREs with an *E* value of 2.3^{-242}, and the other with these sequences with an *E* value of 9.1^{-3}. This may indicate either an alternative recognition pattern recognisable by MmfR orthologues for which binding hasn't yet been observed, or that the *S. hygroscopicus* MmfR orthologue has a different recognition sequence. All putative ARE sequences that have not thus far been investigated would be ideal future screening targets for either the encoded MmfR orthologues of these species or by one of the already characterised MmfR proteins such as MmfR or AvaL1.

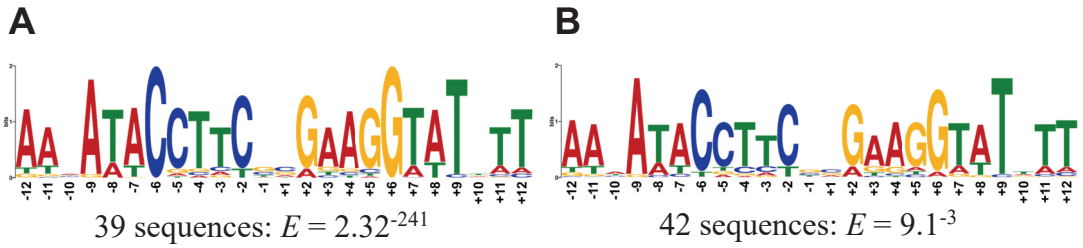


Figure 5-2 – Palindromic MmfR-type ARE motifs from both confirmed and unconfirmed sites across 12 actinomycetes species. (A) Motif from 11 species with strong MmfR-type characteristics. (**B**) Motif from 12 species including weakly characterised sequences predicted in *S. hygroscopicus*. Synthetic sequences bound by AvaL1 are also included.

5.2.3 Comparison of ArpA-subfamily ARE motifs

Although the binding motifs of individual ArpA-subfamily repressors have been published elsewhere, original bioinformatics analyses were performed for alignment between recognised AREs, so that the binding motifs of ArpA-subfamily repressors could be compared against one another. A 22-bp ARE motif (based on the minimum oligonucleotide sequence length required for ArpA binding[204]) for the whole ArpA-subfamily (Figure 5-3A) was first generated using all sequences confirmed to be bound by ArpA-subfamily repressors (Table 5-1), forcing the occurrence of one motif per sequence in the dataset. Using the sequence regions identified from this subfamily motif individual ARE motifs where then constructed for each repressor. As MEME binding motifs need more than one site for motif generation, only those ArpA-subfamily repressors binding two or more AREs were considered. Observations made previously in the literature suggested that gene cluster-specific sites produce more strongly-defined motifs compared with sequences of pleiotropic targets, so individual motifs where built for both gene cluster-specific AREs (Figure 5-3B) and gene cluster-specific + pleiotropic sites (Figure 5-3C).

Table 5-1 – *Streptomyces* ArpA-subfamily repressors for which binding to more than one target has been characterised. More than one site is required for generation of a binding motif in MEME. The number of AREs are indicated for AREs in GBL/AHFCA regulatory cassettes and the primary regulated biosynthetic system + pleiotropic targets, except for ArpA for which only one site *in situ* has been identified, with binding observed to thirty-four synthetic probes. *P* prefix for receptor-types denotes for pseudo. ND = not determined.

Organism	Biosynthetic System	Repressor	Receptor-type	AREs	References
S. griseus	*str* (streptomycin)	ArpA	GBL	1 + 34	(204)
S. coelicolor A3(2)	*mmy* (methylenomycin)	MmfR	AHFCA	3	(139)
		MmyR	*P*-AHFCA	ND	
	cpk (coelimycin)	ScbR	GBL	5 + 12	(66, 76, 98, 99)
		ScbR2	*P*-GBL	7 + 36	
	Pleiotropic regulators	CprA	GBL	ND	(69, 152)
		CprB	GBL	0 + 6	
S. venezuelae	*jad* (jadomycin)	JadR2	*P*-GBL	2	(99, 103)
		JadR3	GBL	4	
	gbn (gaburedin)	SgnR	AHFCA	3	(148)
		GbnR	*P*-AHFCA	ND	
S. avermitilis	*ave* (avermectin)	AvaR1	GBL	6 + 10	(77, 78, 154)
		AvaR2	*P*-GBL	6 + 12	
		AvaR3	*P*-GBL	3	
	azoxy compound	AvaL1	AHFCA	3	This work
		AvaL2	*P*-AHFCA	ND	
S. sclerotialus	*scl* (scleric acid)	SclM1	AHFCA	4	This work
		SclM4	*P*-AHFCA	ND	
S. ambofaciens	*alp* (kinamycin)	AlpZ	GBL	3	(227, 228)
		AlpW	*P*-GBL	3	
S. lavendulae	*far* (indigoidine)	FarA	GBL	5	(107, 229)
		FarR2	*P*-GBL	5	
S. aureofaciens	*aur* (auricin)	SagR	GBL	2	(143)
		Aur1R	*P*-GBL	6	
S. bingchenggensis	*Mil* (milbemycin)	SbbR	AHFCA-like	3 + 11	(79)
S. fradiae	*tyl* (tylosin)	TylP	GBL	3	(114)
		TylQ	*P*-GBL	ND	
S. tsukubaensis	*bul* (tacrolimus)	BulR1	GBL	6	(116)
		BulR2	*P*-GBL	6	
S. pristinaespiralis	*Pap* (pristinamycin)	SpbR	GBL	6	(230, 231)

All generated ARE motifs where significant ($E < 0.05$) apart from JadR2, SagR and SbbR, for which there were either too few sequences or the motifs were poorly defined. The ArpA-subfamily palindromic ARE motif has a consensus of 5'-ACAwACsGsmC|GksCsGTwTGT-3', with the best palindromic match being 5'-ACAAACCGGCC|GGCCGGTTTGT-3'. All of the ARE motifs in Tomtom were significantly matched between their palindromic and nonpalindromic motifs (Table 8-8, Appendix D), with the ArpA-subfamily motif being the most palindromic in both Tomtom ($p = 5.08^{-16}$) and log E value ratio (0.95). Among individual binding motifs the FarA ARE motif was the most significantly palindromic in Tomtom ($p = 1.03^{-14}$), while the SbbR ARE motif was the least ($p = 7.06^{-9}$). From the ratio of log E values, the MmfR-type ARE motif was still the most palindromic (0.92), followed by the AvaR1* (asterisks indicating the motif was built from gene cluster + pleiotropic sites) and TylP motifs where the next most palindromic (log E ratio of 0.88 for both), while the three least palindromic motifs by log E ratios were SbbR (0.29), ScbR2 (0.44) and CprB (0.50). Log E ratios for JadR2, SagR and SbbR could not be accurately determined because the motif E values were not significant enough.

Many of the ARE motifs for these repressors share similar characteristics, as indicated in the ArpA-subfamily motif; The motif is highly degenerate in the central dyad centre region (-3 to +3) and the extreme distal ±9-10 region, whereas the ±9-4 region is generally better conserved, with a higher A/T nucleotide composition. The dyad centre region is less degenerate for both ArpA and MmfR-type AREs motifs, but it is not clear what effect this would have, if any, on ARE recognition. For the SbbR ARE motif A/T nucleotide symmetry is still observed, but the ±9-4 region is much less conserved, and by extension the dyad centre is relatively less degenerate. This is reflected in less significant E values for both palindromic and nonpalindromic motifs. In the ±9-4 region the ±6 C/G nucleotide, as for the MmfR-type ARE, is the most frequently conserved nucleotide across all ArpA-subfamily orthologues but is not fully conserved in all ARE motifs; it is only fully conserved in the gene cluster-specific ARE motifs for BulR1, BulR2, AvaR1, AvaR2, FarA, ScbR, TylP, JadR3, AlpZ, AlpW, and JadR2. Nucleotide ±10 in the TATA…TATA regions (±10-7) has varying compositions, being most commonly occupied in individual ARE motifs by either A/C in the -10 position and T/G in the +10 position. This position is most degenerate in the MmfR-type ARE motif.

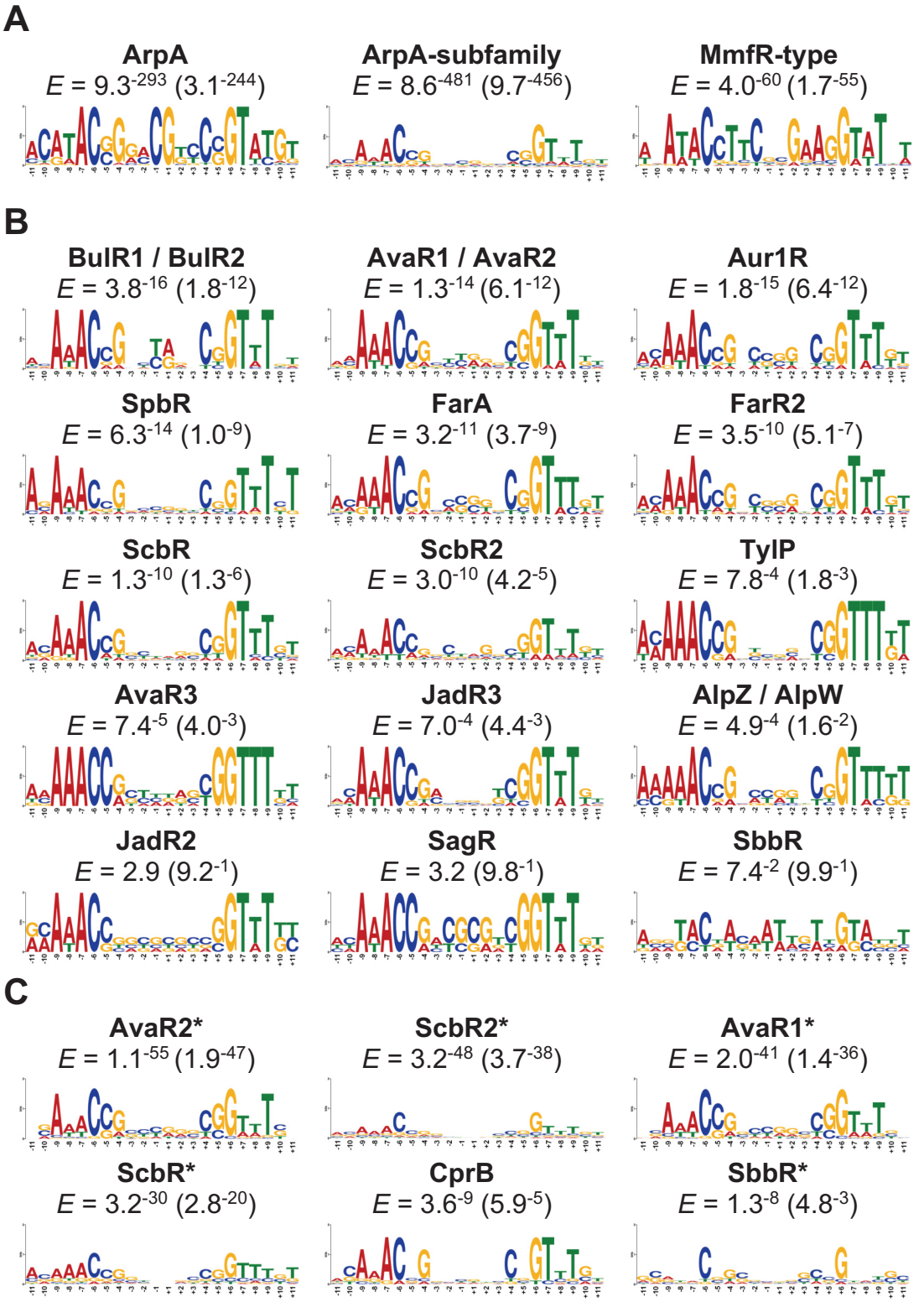


Figure 5-3 – 22-bp palindromic ARE motifs of ArpA-subfamily repressors generated in MEME. (**A**) ArpA-subfamily ARE binding motif generated using all sites bound by an ArpA-subfamily repressor (centre), with the ArpA ARE motif shown on the left and the MmfR-type ARE motif shown on the right. (**B**) Gene cluster specific ARE binding motifs for individual ArpA-subfamily repressors. (**C**) Binding motifs for ArpA-subfamily repressors shown to function as pleiotropic regulators, with motifs generated from both gene cluster-specific sites and pleiotropic targets (motifs differentiated with asterisks). *E* values for nonpalindromic motifs are given with *E* values of palindromic motifs in parenthesis; nonpalindromic AREs motifs are not shown. All motifs are calculated with a forced occurrence of one motif per sequence within each dataset, as has been shown experimentally (references in Table 5-1). All motifs are sorted in order of increasing palindromic *E* values.

5.2.4 Characterising pleiotropic AREs of GBL/Pseudo-GBL receptors

To date there have only been five ArpA-subfamily repressors for which gene cluster-situated repressors have demonstrable binding to regulatory sites outside of their primary regulated gene clusters; ScbR/ScbR2[76], AvaR1/AvaR2[77, 78, 154] and SbbR[79]. For AREs bound within gene clusters there isn't much difference in the ARE binding motifs of either GBL or Pseudo-GBLs working in tandem, with the same sites being bound by both repressors in some systems (AvaR1/AvaR2, BulR1/BulR2, AlpZ/AlpW). However, when comparing the ARE motifs of GBL and pseudo-GBL repressors by factoring in their pleiotropic targets (Figure 5-3C) the differences are more apparent; compared with the gene cluster-specific ARE motifs the pleiotropic ARE motifs are much more degenerate, though less significantly for the AvaR1 and AvaR2 systems. The ARE motif of ScbR2 is also more degenerate compared to ScbR, and in both AvaR1/AvaR2 and ScbR/ScbR2 systems the pseudo-GBL receptor binds more pleiotropic target AREs than their GBL receptor counterpart. This is reflected in % GC content analysis (Table 8-8, Appendix D); the mean % GC content of gene cluster-specific AREs (for which there is no significant difference in % GC content between all GBL and pseudo-GBL receptor AREs) for these five repressors is 48.0 % ± 9.1 %, with the maximum % GC ARE being 64% GC by content (bound by AvaR1/AvaR2). Factoring in the pleiotropic ARE targets increases the % GC content to 58.5 % ± 12.6 %, with the putative SbbR site *sbi_05779* (5'-GCagCCGGCcC|GcGCCGGagGC-3') having 91 % GC content.

The issue with characterising these pleiotropic sites is that they are putative, based on bioinformatic predictions using the gene cluster-specific ARE motifs. These binding sites were confirmed using gel shift assays, but these are performed using PCR-amplified probes much larger than synthetic probes such as those designed in for this work, usually hundreds of base-pairs long. In the case of AvaR1/AvaR2 these pleiotropic AREs are more similar to their gene cluster counterparts AREs, but this is not the case for ScbR, ScbR2 and SbbR. Therefore, it is difficult to perform an accurate analysis of ARE motifs bound pleiotropically. The problem may be in our understanding of ArpA-subfamily binding motifs; Li, X. *et al* have deposited an alternative hypothesis for the mode of DNA recognition by ScbR and ScbR2[76]. From their own bioinformatic analysis they observed that these pleiotropic sites contained palindromic AT-rich regions which were not always equidistant from the proposed dyad centre, and that these AT-rich regions were sometimes found only occurring once in the target sequence; a putative half-dyad region (or half-ARE). In the sites unique to ScbR2 single occurrences of these half-dyad regions were found more frequent than observed in the sites shared by both ScbR and ScbR2. Thus, they interpreted the binding motifs as 10-bp half-dyad sequences, with consensus sequences of msGyTTsTTd for ScbR and dyTysTysws for ScbR2.

MEME was used to identify these shorter half-dyads in these pleiotropic targets for all the ArpA-subfamily pleiotropic regulators, again forcing one occurrence per site (Figure 5-4). These half-dyad regions were better defined than the 22-bp palindromic ARE motifs described previously, although again for SbbR the motif was poorly defined. This may indicate that only limited palindromacy around the dyad centre may be required if a sufficiently recognisable half-dyad region is present in the ARE sequence. In Chapter 4.4.4 it was speculated that positive cooperativity for binding of CprB dimers to the *OPB* ARE is observed because of the poor palindromacy around the dyad centre (5'-gcaggCgGcaC|GgtCtGttgag-3') necessitating the initial binding of a single dimer to distort DNA, allowing for sequential binding of the second CprB dimer. The *OPB* sequence could be interpreted as containing only a single half-dyad ARE-like region (5'-gcaggCgGcaC|**GgtCtGttgag**-3'), which serves as an anchoring point for the first dimer. By anchoring to this region, the DNA is subsequently distorted, allowing access to the major groove in the other previously unrecognisable half-dyad region by the second dimer.

It is possible that a similar scenario occurs also for other ArpA-subfamily repressors that can bind these proposed half-ARE sites. With regards to apparent variable distances between symmetrical half-dyad regions, it is not likely that each half-dyad region serves as a single binding site for these repressors. From what is known from the crystal structures of CprB and other TFRs in complex with DNA conformational the two dimers bound by ArpA-subfamily repressors have to be in close proximity to one another for binding to occur, as distortion of DNA by a single dimer binding is unlikely to induce the required DNA confirmation at a distant region to make the site available for the second dimer. Rather, these distal sites may serve instead as independent half-dyad anchoring points result in two possible complex formations in these bound regions.

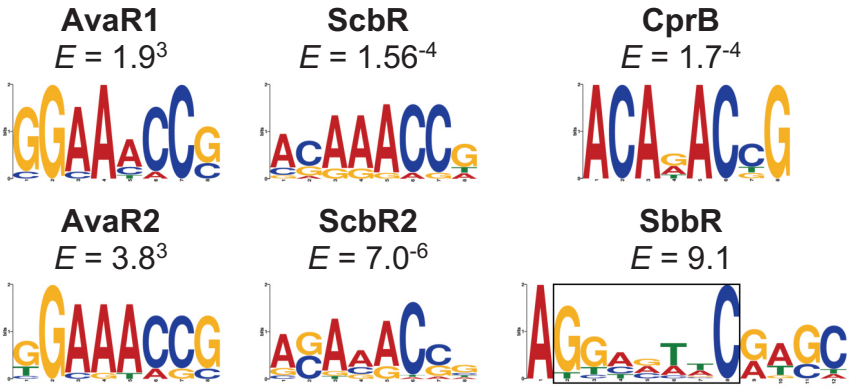


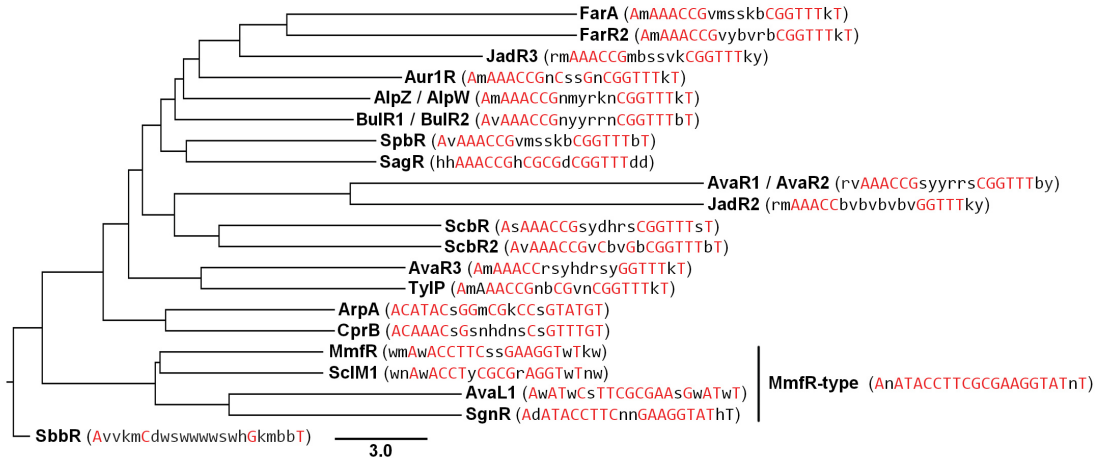
Figure 5-4 – Proposed half-dyad ARE motifs identified in the pleiotropic target regions of *S. avermitilis* AvaR1/AvaR2, *S. coelicolor* A3(2) ScbR/ScbR2/CprB and *S. bingchenggensis* SbbR. Half-dyad regions indicated correspond to the –11 to -4 dyad region of ARE motifs in Figure 5-3.

5.2.5 Comparative analysis of ArpA-subfamily and MmfR-type ARE motifs

To compare the similarity between gene cluster-specific ARE motifs, an unrooted Unweighted Pair Group Method with arithmetic Mean (UPGMA) tree was generated using the PHYLogeny Inference Package (PHYLIP), with pairwise comparative data of ARE motifs generated in Tomtom. The tree (Figure 5-5A) shows that the MmfR-type ARE motifs form a clade distinct from the other ArpA-subfamily ARE motifs but have more in common with them than the SbbR ARE motif. There is no significant differential branching between GBL and pseudo-GBL gene cluster-specific AREs either, with different ARE motifs of repressors from the same gene cluster sometimes clustering together in a single clade (FarA/FarR2 and ScbR/ScbR2). To determine what differentiates the MmfR-type ARE motif, the consensus sequence of each ARE motif was compared against the consensus sequence for the global ArpA-subfamily ARE motif (Figure 5-5B). CprB and ArpA are the strongest match to this consensus sequence (these two repressors also cluster together in the tree), with all AHFCA receptor ARE motifs and the SbbR ARE motif scoring relatively poorly against the ArpA-subfamily consensus.

The most disparate nucleotides positions in the MmfR-type ARE motifs are positions ±3 and ±4, which are occupied by A/T nucleotides that are comparatively rare in the other ARE motifs. These positions correspond with the ±2 A/T dinucleotide block observed in the dinucleotide block consensus sequence (5'-WWWS**W**S|S**W**SWWW-3'), and analysis clearly shows that this dinucleotide block is overwhelmingly occupied by C/G nucleotides in other GBL/Pseudo-GBL receptor ARE motifs (Figure 8-13C, Appendix D), forming the dinucleotide block consensus sequence 5'-WWWS**S**S|S**S**SWWW-3'. This is also reflected in % GC content analysis, which shows that the mean % GC content of the 13 sequences bound by AHFCA receptors is 36.5 ± 6.8 %; the lowest of any of the other characterised ArpA-subfamily repressors (Table 8-8, Appendix D). It is not clear how necessary this A/T dinucleotide block is for AHFCA receptor binding (or alternatively affecting the binding of other ArpA-subfamily repressors), but the *sclM4/sclM5*$_{ARE}$ sequence is the only ARE lack this at least one A/T dinucleotide block in one half-dyad.

A



B

Motif	Consensus Sequence	Score
ArpA-subfamily	ACAWACSGSMCGKSCSGTWTGT	-
ArpA	ACATACSGGMCGKCCSGTATGT	1
CprB	ACAAACSGSNHDNSCSGTTTGT	0.89
SpbR	AVAAACCGVMSSKBCGGTTTBT	0.88
FarA	AMAAACCGVMSSKBCGGTTTKT	0.88
AlpZ/AlpW	AMAAACCGNMYRKNCGGTTTKT	0.86
Aur1R	AMAAACCGNCSSGNCGGTTTMT	0.86
TylP	AMAAACCGNBCGVNCGGTTTKT	0.85
ScbR2	AVAAACCGVCBVGBCGGTTTBT	0.85
ScbR	ASAAACCGSYDHRSCGGTTTST	0.82
FarR2	AMAAACCGVYBVRBCGGTTTKT	0.82
AvaR3	AMAAACCRSYHDRSYGGTTTKT	0.80
AvaR1/AvaR2	RVAAACCGSYYRRSCGGTTTBY	0.80
BulR1/BulR2	AVAAACCGNYYRRNCGGTTTBT	0.80
JadR3	RMAAACCGMBSSVKCGGTTTKY	0.76
SagR	HHAAACCGHCGCGDCGGTTTDD	0.73
JadR2	RMAAACCBVBVBVBVGGTTTKY	0.70
MmfR	WMAWACCTTCSSGAAGGTWTKW	**0.68**
SclM1	WNAWACCTYCGCGRAGGTWTNW	**0.66**
SgnR	ADATACCTTCNNGAAGGTATHT	**0.66**
AvaL1	AWATWCSTTCGCGAASGWATWT	**0.59**
SbbR	AVVKMCDWSWWWWSWHGKMBBT	0.50
AvaR1*	VVAWACCGSCCGGSCGGTWTBB	0.88
ScbR*	AVAAACCGSBNNVSCGGTTTBT	0.82
ScbR2*	AVAAACSGSBNNVSCSGTTTBT	0.81
AvaR2*	NSAAACCGSYYRRSCGGTTTSN	0.80
SbbR*	GSVBVCSGSSMKSSCSGBVBSC	0.61

Figure 5-5 – Bioinformatic comparative analysis of palindromic ARE motifs of ArpA-subfamily repressors. (**A**) Unrooted unweighted pair group method with arithmetic mean (UPGMA) Tree of ArpA-subfamily binding motifs using AREs found within primary biosynthetic gene clusters. Binding motifs were scored pairwise in Tomtom, then using the log *p* values the tree constructed using PHYLIP. Pleiotropic regulatory motifs were excluded, except for the pleiotropic regulator CprB. (**B**) Scoring of consensus motifs for individual repressors against the ArpA-subfamily consensus motif. Each position was given a score based on probability of a nucleotide match to the ArpA-subfamily consensus and then averaged across the sequence. An example being D (A/G/T) scored against W (A/T) having a score of 0.67, or a T against W having a score of 1. Nucleotides matching the consensus are highlighted in **black**, while nucleotides with a partial match are highlighted in **grey**. Scores for the MmfR-type repressors are indicated in bold. Shown below this are the pleiotropic ARE consensus motifs (differentiated with asterisks).

Nucleotide IUPAC codes for are as follows:

R = A/G	Y = C/T	S = G/C
W = A/T	K = G/T	M = A/C
B = C/G/T	D = A/G/T	H = A/C/T
V = A/C/G	N = any base	

5.2.6 Predicting targets of AHFCA/pseudo-AHFCAs receptors across the genomes of *S. coelicolor* A3(2) and *S. avermitilis*

5.2.6.1 Initial ARE predictions

It is hypothesised that the AHFCA/Pseudo-AHFCA and GBL/Pseudo-GBL regulatory systems in *Streptomyces* spp. are analogous in their role in regulating both their gene cluster-specific pathways as well as coordinating pathways through pleiotropic regulation. ARE predictions across the *S. avermitilis* genome have previously been performed by Dr Poon[112] using a custom automated script to extract regions 300-bp upstream and 100-bp downstream of all open reading frame (ORF) start codons, and then parsing the sequences through FIMO using a motif generated from the *S. coelicolor* A3(2) MmfR AREs, from which 58 potential AREs were identified. Using new bioinformatic analyses 1,019 and 1,794 unique predictions were made within the *S. coelicolor* A3(2) and *S. avermitilis* genomes respectively (available in the Electronic Supplementary Materials) using FIMO with a *p* value cut-off of $p < 10^{-5}$, with MEME nonpalindromic ARE motifs for each species-specific AHFCA receptor used as a search template. In *S. coelicolor* A3(2) 446 sequences (43.7 %) were identified at a distance of 300-bp to -26-bp from ORF start codons, and 135 (13.6 %) were within 300-bp to 500-bp upstream and between 26-bp to 126-bp downstream of ORF start codons. In *S. avermitilis* for the same parameters there were 736 (41.0 %) and 312 (17.4 %) predicted AREs respectively (including redundant predictions from Dr Poon's dataset).

In both genomes there were no sequences which scored better than the three AREs bound within the AHFCA-regulated gene clusters, and the confirmed ARE sequences scored significantly better than other predictions; for example, $mmyR_{ARE}$ has a *p* value of 5.78^{-15}, and the next most significant prediction was *SCO5240* (*whiB*) with a *p* value of 1.73^{-8}. Some of the strongest predictions based on *p* value were also found within ORF sequences, one example being the 5'-AATaGAcCTTCt|tGAAGaTCcATT-3' found 889-bp downstream of the start codon for the 1,431-bp gene *saverm_1645,* and 605-bp upstream of *saverm_1644*. It is possible that the repressors bind these sites, but the biological effect is difficult to predict.

5.2.6.2 Prediction of other AREs within AHFCA-regulated gene clusters

It is possible that there are other genes besides the AHFCA regulatory cassette and transcriptional activator which are regulated by the AHFCA/Pseudo-AHFCA receptors within the AHFCA-regulated gene clusters of *S. coelicolor* A3(2) and *S. avermitilis*. Half-ARE-like sequences have previously been identified upstream of the ORFs of *mmyQ*, *mmyJ*, *mmr* and

mmyT within the *mmy* gene cluster, and have been proposed as potential targets for MmyR[67, 112]. Potential ARE targets within these gene clusters were extracted from the FIMO predictions (Table 5-2). An ARE was predicted upstream of *mmyQ* which overlaps with the previously reported half-ARE-like sequence, but no AREs were predicted in the regions of the other half-ARE-like sequences. Two new AREs predicted within the *mmy* gene cluster include one within the ORF of *mmyX*, and another upstream of *mmyR*, offset 4-bp from the MmfR binding site. Within the *S. avermitilis* AHFCA-regulated gene cluster a further nine possible AREs were predicted; an ARE within the overlapping region of the *avaB*/*avaC* operon, an ARE upstream of *saverm_2278*, an ARE overlapping the *fabF* start codon, an ARE within the ORF of *fabI2*, two AREs upstream of *saverm_2300*, and two additional ARE upstream of *saverm_2301*, the first of which is offset 3-bp from the AvaL1 binding site. Many of these putative AREs, as for the *mmy* gene cluster, may contain a half-ARE like sequence.

Table 5-2 – Putative AREs within the AHFCA-regulated gene clusters of *S. coelicolor* A3(2) and *S. avermitilis*. Palindromic nucleotides around the dyad centre are underlined. Confirmed AHFCA receptor AREs are in **bold**. Distance is indicated as the shortest distance from the ORF start codon. Each sense strand for the predicted AREs where compared against a 24-bp AHFCA receptor half-ARE motif (5'-AAnATACCTTCGn$_{12}$-3') using MAST; half-dyads for which $E < 1^{-4}$ are highlighted.

ARE sequence (5' to 3')	Regulated target	*p*	Distance
GTCAAACCTTCC|CGAAGGTTTTTT	***mmyB***	**4.22^{-15}**	**13**
TGCATACCTGCG|GAAAGCCAGGAA	*mmyQ*	8.24^{-06}	21
ACGAAACCTCGG|GGCAGGTGGCTC	*mmyX*	3.86^{-05}	-209
AATATACCTGCG|GGAAGGTATTAT	***mmfR/mmfLHP***	**4.11^{-16}**	**76/94**
AACATACCCTCG|GGAAGGTATGTT	***mmyR***	**5.78^{-15}**	**44**
TACCTTCCCGAG|GGTATGTTTTCC	*mmyR*	3.49^{-05}	40
AAAAGAGCCTCT|CGCCCTCCTACT	*avaC*	8.74^{-05}	-49
AGAATTCGTTCA|CGAACGTATCTT	***avaBC/avaL2***	**1.41^{-14}**	**25/19**
AATATACCTGCG|CGAAGGTATATT	***avaA/avaL1***	**2.25^{-14}**	**72/95**
CATCTACCTGTG|CGACCGCATGAA	*saverm_2278*	2.01^{-05}	336
AGGAGCGGAGCG|CGAAGGTCTGAT	*fabF*	3.95^{-05}	-2
ACACTACGGTCT|CGACGGGGTGGT	*fabI2*	5.77^{-05}	-260
GAAACAGCTTCC|GAAAGGGGTACG	*saverm_2300*	3.21^{-05}	3
AGGGGAAGACCG|GAAAGGAATTCA	*saverm_2300*	9.71^{-05}	60
ATAATTCCTTTG|AGAAGGTATTTT	***saverm_2301***	**6.53^{-16}**	**178**
AGAATAATTCCT|TTGAGAAGGTAT	*saverm_2301*	7.42^{-05}	175
CATACATCCTCG|CACAGAATAATT	*saverm_2301*	8.23^{-05}	160
ACTCGGCCTGCT|GGAACGGCTGTT	*fadE10*	7.49^{-05}	250

5.2.6.3 Mapping putative MmfR-type AREs to regions bound by GBL/pseudo-GBL receptors

As has been described previously in this chapter, a single ARE site can often be bound by multiple GBL/pseudo-GBL receptors. This is demonstrated in GBL/Pseudo-GBL receptors working in tandem to regulate the gene clusters in which their genes are encoded but can also be seen by the binding ArpA-subfamily repressors within different regulatory systems, such as the binding of AREs in the *cpk* gene cluster by both the gene cluster-specific regulators ScbR/ScbR2 as well as the pleiotropic regulator CprB in *S. coelicolor* A3(2). Given that the

MmfR-type ARE motif shares some similarity with the ARE motifs of other ArpA-subfamily repressors, it is possible that cooccurring AHFCA and GBL systems regulate shared targets.

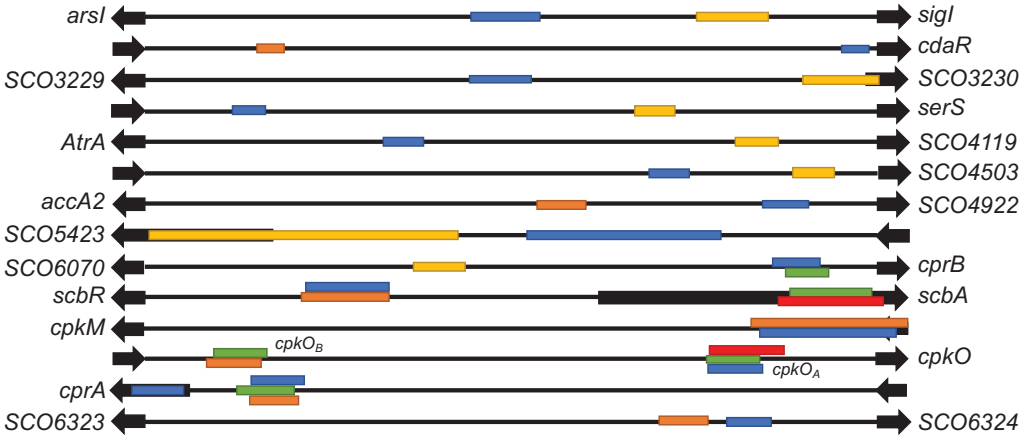
The FIMO predictions were cross-referenced with known binding sites of GBL and pseudo-GBL receptors in *S. coelicolor* A3(2) and *S. avermitilis* (Figure 5-6)[69, 76-78]. In *S. coelicolor* A3(2) putative MmfR AREs can be found within the *cpk* gene cluster at the *scbR*/*scbA* divergent promoter site bound by ScbR/ScbR2, at the *cpkM* site recognised by ScbR/ScbR2, and at the $cpkO_A$ ARE bound by ScbR/CprB. Another MmfR ARE was predicted to overlap *OPB* upstream of *cprB* (bound by CprB), downstream of a different ARE bound by ScbR2. Two MmfR AREs were also predicted in the regulatory region of *cprA*; one upstream of *cprA* overlapping a site bound by ScbR/ScbR2/CprB, and one within the ORF of *cprA*. Finding predicted MmfR AREs in the upstream regions of these targets is not surprising, as these sites are amongst those closest conserved to the ArpA type recognition motif.

For the sites pleiotropically regulated by other ArpA-subfamily repressors, MmfR AREs are predicted in the upstream regions of *arsI/sigI*, *SCO3229/SCO3230*, *serS*, *SCO4118/SCO4119*, *SCO4503*, and *cprB*, which are bound by ScbR2, and upstream of *cdaR*, *accA2/4922*, *5423*, *cprA*, and *SCO6323/6324*, which are bound by both ScbR and ScbR2. However, none of these MmfR ARE predictions overlap with the putative binding sites of ScbR/ScbR2. The putative MmfR AREs are downstream of the putative ScbR/ScbR2 binding sites in the intergenic regions of *arsI*, *cdaR*, *SCO3229*, *SCO4118*, *SCO4922* and *SCO6324*, and if binding is confirmed may indicate coregulation of these targets. Conversely, MmfR AREs are predicted upstream of ScbR2 binding sites in the intergenic regions of *sigI*, *SCO3230*, *serS*, *SCO4119*, *SCO4503*, *accA2* and *SCO5423*, and thus are less likely to regulate these targets.

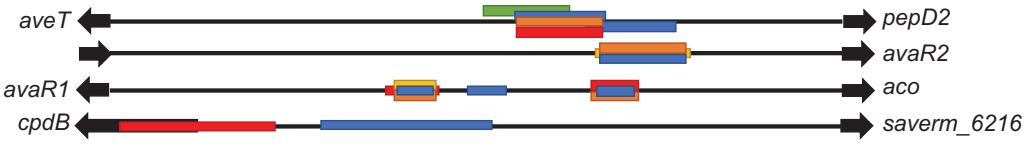
In *S. avermitilis* there are fewer AvaL1 ARE predictions found in the region of known AvaR1/AvaR/AvaR3 binding sites. There are three AvaL1 AREs predicted in the upstream region of *avaR1*/*aco*; one overlaps the *avaR1* promoter site bound by AvaR1/AvaR2/AvaR3, and another overlaps the *aco* site bound by AvaR1/AvaR2. Similarly, a site is predicted upstream of *avaR2* which is also bound by AvaR1/AvaR2. Three other sites are located in upstream regions of genes pleiotropically regulated by AvaR1/AvaR2; two are located in the intergenic region of the TFR-encoding gene *aveT* (positive avermectin biosynthesis regulator[232]) and *pepD2*, one of these sites overlapping the AREs that are predicted to be bound by AvaR1 and AvaR2. These sites also overlap with the pseudo-palindromic binding sequence of AveT itself, which regulates its own expression as well as *pepD2*. The other site is found upstream of the putative AvaR1 binding site overlapping the start codon of *cpdB*.

A

S. coelicolor A3(2) sites



S. avermitilis sites



B

ARE sequence (5' to 3')	Regulated target	*p*	Distance
AACACACTTGCG | GTCAGCTCGGTA	*arsI/sigI*	4.41^{-05}	112/116
GATACCCCCGCG | TCAAGTTGTGCT	*cdaR*	4.41^{-05}	8
AAGATAACCGCA | GGTAGGAGCCTT	*Sco3229/3230*	4.46^{-05}	118/126
CATTGACCCGCC | GAAGGGGATGTG	*SCO3961(serS)*	9.85^{-05}	61
AATTGACCCCCG | CTCAACTATGTT	*SCO4118(atrA)/4119*	2.60^{-05}	138/260
GACTGACCGTCG | GTAACCCATGTC	*Sco4503*	5.20^{-05}	104
AACGTACTTGGG | GTAACAGCTGTC	*SCO4921(accA2)/SCO4922*	5.52^{-05}	316/34
AATATTCCGTCT | GGCGGAAAATAC	*SCO5423*	4.33^{-05}	23
TAAAGGCAGGCG | GCACGGTCTGTT	***SCO6070/CprB***	**4.51^{-05}**	**304/27**
AAGATACAGACT | GAGCGGTTTTTT	***scbA***	**1.61^{-05}**	**40/53**
TGAAAACCGGCG | AGGAGGTTTGTT	***cpkM***	**1.82^{-07}**	**104**
GACAAACCGGTG | TGCTGGTTTGTA	***cpkO$_A$***	**1.68^{-05}**	**46**
AACGGACCCGCG | CGACGATTGTCC	*cprA*	3.00^{-05}	-22
AACAGACCGTGT | GCCTGTTTTTAC	***cprA***	**3.33^{-05}**	**27**
AAGATAGTTCTG | GCTACGTTATTT	*SCO6323/SCO6324*	7.06^{-05}	306/55
ACGACACGTACG | AAACGGTTTCGT	*aveT/pepD2*	4.23^{-05}	102/60
AAAACATGTGCG | TGACCCACGACA	*aveT/pepD2*	2.35^{-05}	120/42
AAGAAACCAGCA | TAGCGGTTTGTT	***avaR2***	**1.32^{-05}**	**41**
ATAAAACCAGCT | GAACGGTTTTTC	***avaR1/aco***	**1.16^{-05}**	**170/250**
GATATGTGTTTT | TGAGTGGTTTGT	*avaR1/aco*	8.46^{-05}	214/206
AAGATACGTACT | AGACGGTTTTGT	***avaR1/aco***	**4.28^{-08}**	**292/128**
AATCTACGCGCG | TCATTGGCCTTT	*cpdB/saverm_6216*	4.92^{-05}	17/48

Figure 5-6 – Putative AHFCA receptor AREs in the binding regions confirmed to be bound by other ArpA-subfamily repressors in *S. coelicolor* A3(2) and *S. avermitilis*. (**A**) Annotated binding regions with binding sites coloured by ArpA-subfamily repressors which bind it; colours for *S. coelicolor* A3(2) sites are **ScbR**, **ScbR2**, **ScbR/ScbR2**, **MmfR** (putative) and **CprB**. Colours of ArpA-subfamily repressor binding sites in *S. avermitilis* are **AvaR1**, **AvaR2**, **AvaR3**, **AveT** and **predicted AvaL1**. AvaR1/AvaR2 binding sites not upstream of *avaR1*, *aco* or *avaR2* are predicted based on bioinformatics of regions demonstrated to be bound by AvaR1/AvaR2 *via* gel shift assays[77, 78]. (**B**) Putative ARE sequences. Palindromic nucleotides around the dyad centre are underlined. Putative AHFCA receptor AREs overlapping with GBL/Pseudo-GBL AREs (putative or otherwise) are in **bold**. Distance is indicated as the shortest distance from the ORF start codon. Each sense strand for the predicted AREs where compared against a 24-bp AHFCA receptor half-ARE motif (5'-AAnATACCTTCGn_{12}-3') using MAST; half-dyads for which $E < 1^{-4}$ are highlighted.

5.2.6.4 Refining and testing binding for these predictions

Assaying the whole set of predicted AREs using a method such as SPR is not practical, as it is not designed for screening so many targets. Microarray-based methods such as ChIP-on-chip would be much better suited for screening the binding of AHFCA receptors to these synthetic probes. However, some of the more significant predictions can be tested on SPR to determine if there is a basis in the hypothesis of pleiotropic regulation by AHFCA receptors and could also be used to test for the binding pseudo-AHFCA receptors, which are proposed to bind similar sequences based on homology. To identify the strongest candidates for screening, each sequence was assessed based on two criteria; the first being that *p* value of the FIMO prediction must be less than 1^{-6}. Secondly, all predicted AREs with a *p* value < 1^{-4} were compared against the palindromic MmfR-type ARE motif in MAST in two motif configurations; one the full palindrome motif (5'-AANATACCTTCGCGAAGGTATATT-3') and the other a half palindrome (5'-AANATACCTTCGn_{12}-3'), and only those that scored better than *sclM4/sclM5*$_{ARE}$ (which was the lowest scoring in the construction of the initial motif) were selected. These refined predictions are shown in Table 5-3 and Table 5-4.

SPR experiments were performed to test the binding AvaL1 and AvaL2 to AREs predicted previously across the *S. avermitilis* genome by Dr Poon[112], utilising immobilisation of AvaL1 and AvaL2 to NTA-chips. However, an oversight regarding oligonucleotide design for the ARE probes was made, so the validity of the results cannot be attested for. The results of these experiments remain available in the Electronic Supplementary Materials. However, they may suggest two novel targets for AvaL2 (which requires future confirmation), and at the very least shows demonstrable biological activity of the recombinant AvaL2 repressor binding to DNA.

Table 5-3 – Most significant AHFCA receptor AREs predicted using FIMO in *S. coelicolor* A3(2). Most significant sequences in FIMO are indicated if $p < 1^{-6}$. Pm and ½Pm indicate sequence matches to the palindromic (5'-AANATACCTTCGCGAAGGTATATT-3') and ½ palindromic (5'-AANATACCTTCGn_{12}-3') AHFCA receptor motifs that score in MAST at least as well as or better than the worst scoring confirmed ARE *sclM4/sclM5*$_{ARE}$. Confirmed AHFCA receptor AREs are in **bold**. Distance is indicated as the shortest distance from the ORF start codon. Half-dyad matches to scores for which $E < 1^{-4}$ are highlighted.

$p < 1^{-6}$	Pm	½Pm	ARE Sequence (5' to 3')	*S. coelicolor* A3(2) Genes	Distance
		✓	ATGAGACCTACG | GGAAGATGAGTT	*SCO0026*	-358
✓			CTCATGTCTTCG | AGAAGGTAGGAC	*SCO0171*	-47
	✓		AACAAACTCGCA | ATGAAGTTTTTT	*SCO1105/SCO1106*	156/100
	✓	✓	AAGGACCCGTCC | GGGAGGTACGAC	*SCO1200/SCO1201*	107/-30
✓			AACATGCCTACC | GGTCGGTATGCT	*SCO1711*	41
		✓	AAGAAAAATTGC | GCAATGGATGTT	*SCO2047*	20
	✓	✓	AAGGAACCCGCG | GGACGGGAATTG	*SCO2951*	211
	✓		AGGATTCCGGCC | GGACGGTATGCC	*SCO3119*	109
		✓	AAGAAGGCGGGC | GGAAGGTCTGAC	*SCO3156*	16
		✓	ACCAAACGCTCC | GCAACCTACGAC	*SCO3190*	34
		✓	AGAAGACCTTCG | GCACCGGCTGAC	*SCO3632*	-436
	✓	✓	AAAAAACCTTCA | GGCAACCACAAC	*SCO3731*	73
		✓	AACATGAGTACG | GCAATGTATGTA	*SCO3775/SCO3776*	146/-168
	✓	✓	AGAACACCGGCA | TGAAGGTCAGAC	*SCO4265/SCO4266*	390/-1156
		✓	AACGTACCGTCG | GGAAGGTCGCCG	*SCO5023/SCO5024*	-935/168
✓			GATGTCCCTGCG | CGAAGGGGTGAT	*SCO5088/SCO5089*	-1037/187
✓	✓	✓	AAAAAACTTTCC | GGAAGGGGTTGT	*SCO5240 (whiB)*	203
	✓	✓	GGCAAACAGGCG | GGAAAGCATGAA	*SCO5254 (sodN)/SCO5255*	-21|139
✓			GTACTACCTGCG | GGAAGGGCTGTT	*SCO5550/SCO5551*	-53/134
✓	✓	✓	TGAAAACCGGCG | AGGAGGTTTGTT	*SCO6268 (cpkM)*	104
	✓	✓	AAGAAGTCCGCG | AAAAGGTGTAAC	*SCO7105/SCO7106*	174|204
✓			AATATCCCCGTG | GCACGGTGTGTC	*SCO7187*	68
✓			AACAACCGTGCG | ACAAGGGCTTTT	*SCO7762*	-579
		✓	GTAAAACGTTCG | GGGTGGTGTCAC	*SCP1.153*	103
		✓	TGCATACCTGCG | GAAAGCCAGGAA	*SCP.231 (mmyQ)*	21
✓	✓	✓	**GTCAAACCTTCC | CGAAGGTTTTTT**	***mmyB*$_{ARE}$**	**13**
✓	✓	✓	**AATATACCTGCG | GGAAGGTATTAT**	***mmfR/mmfLHP*$_{ARE}$**	**76/94**
✓	✓	✓	**AACATACCCTCG | GGAAGGTATGTT**	***mmyR*$_{ARE}$**	**44**

Table 5-4 – **Most significant AHFCA receptor AREs predicted using FIMO in *S. avermitilis*.** Most significant sequences in FIMO are indicated if $p < 1^{-6}$. Pm and ½Pm indicate sequence matches to the palindromic (5'-AANATACCTTCGCGAAGGTATATT-3') and ½ palindromic (5'-AANATACCTTCGn$_{12}$-3') AHFCA receptor motifs that score in MAST at least as well as or better than the worst scoring bound ARE *sclM4/sclM5*$_{ARE}$. Confirmed AHFCA receptor AREs are in **bold**. Distance is indicated as the shortest distance from the ORF start codon. Half-dyad matches to scores for which $E < 1^{-4}$ are highlighted.

$p < 1^{-6}$	P	hP	ARE Sequence (5' to 3')	*S. avermitilis* Genes	Distance
	✓	✓	AAGAGACCATCG | CAGAGGTCGAAC	*saverm_28*	-730
		✓	CTACTACCTGCT | GGAAGGCATCTT	*saverm_44*	-1004
		✓	AGAATTCATTCT | GAAATGATCAGT	*saverm_RS39100/saverm_88*	*convergent*
		✓	CAGAGATGTCCG | CGAACGAGTACA	*saverm_117*	-322
		✓	CAAGTACCTGCG | CAACGGAGACCT	*saverm_563 (mhpA)*	75
		✓	ATTCCACTCTGT | CGAAGGAATTAG	*saverm_889/saverm_890*	335/52
		✓	AAGACGCGACGT | TGAAGGTCTATT	*saverm_974/saverm_RS39785*	-721/166
	✓	✓	AAGAAACTTGCG | AAACGGCTTTAA	*saverm_995/saverm_996*	377/-41
✓		✓	AAAATAACGGCT | CAAATTAGCTAT	*saverm_1037 (zmp1)*	118
✓			AGATGAGCTTCT | CGAAGGTGTTCT	*saverm_1573/saverm_1575*	-1234/1275
		✓	AAGAGACGTCCG | CAGACGAGGAGT	*saverm_1594*	420
✓		✓	AATAGACCTTCT | TGAAGATCCATT	*saverm_1644/saverm_1655*	605/-889
✓			GAGATCCGCTCG | TGAAGGACGTCT	*saverm_1702 (pca2)*	-13
		✓	AAAATTCTTGCT | GTATTGGATCTT	*saverm_2262 (pkn5)*	73
✓	✓	✓	**AGAATTCGTTCA | CGAACGTATCTT**	***avaL2/avaBC*$_{ARE}$**	**19/25**
✓	✓	✓	**AATATACCTGCG | CGAAGGTATATT**	***avaL1/avaA*$_{ARE}$**	**95/72**
✓	✓	✓	**ATAATTCCTTTG | AGAAGGTATTTT**	***saverm_2301*$_{ARE}$**	**178**
✓			AAAGTATGTGCG | GAATGGTGTCTT	*saverm_2418/saverm_2419*	convergent
✓			AAGAGATGTGGT | CCAACGACTCAT	*saverm_2587*	-39
✓			CATCCACCTGCT | CGAACGACTTCT	*saverm_2744/saverm_2745*	286/-1192
		✓	ACAATCACTGAT | TGTAGGAACATA	*saverm_2821*	43
		✓	AGTACAGCCTAT | CGAAGGAAGGTT	*saverm_2916 (thrB)*	126
	✓		AAGCACGCTGCT | TAAGCGTTTTCT	*saverm_2928/saverm_RS40080*	448/-1432
		✓	AAAAGACGTCTC | CTACGGAGTAAT	*saverm_2987/saverm_2988 (sodN)*	82/39
	✓	✓	AAAAAACTTTCC | GGAAGGGGTTGT	*saverm_3016 (wblE)*	202
		✓	AGGAGACGTACG | CGAAGGCGTACG	*saverm_3037/saverm_3038*	460/-361
		✓	GAGGTGCGATGT | CACATGAGTAAT	*saverm_3196*	-21
✓	✓	✓	CGAAAAAGTACA | TAAACGTATATT	*saverm_3450*	56/112
✓	✓		AAGATACGTACT | AGACGGTTTTGT	*saverm_3705 (avaR1)/saverm_3706 (aco)*	292/128
		✓	AAAGTTCGGTAA | TACAGGAATTGT	*saverm_RS40200/saverm_3722*	-239/452
✓			CATATTCCAGCG | CGAAGGAGTACG	*saverm_3837*	168
		✓	TAGTTCCCAGCG | CGAAGGAATGGT	*saverm_3907 (cspB1)*	126
	✓	✓	AAAACACCTTCG | CGTATGACCCCT	*saverm_3994*	-411/675
		✓	CAGAAACCTCCG | CAAACGGGTACG	*saverm_4041/saverm_4042*	-66/358
	✓		AAGACACGTAAG | AAAAGGAAGCCC	*saverm_4190*	0
✓			TTGCGACCTGCG | CAAACGTGTTCT	*Trn51*	112
		✓	GTGAGAGCGTCT | CGAAGGAGTATT	*saverm_4321*	-661
✓	✓	✓	GAGATACGTGCG | CAGACGCAGGAT	*saverm_4376/saverm_4377*	1364/1235
		✓	ATTAAACCTGCT | TAATGCGCTGCT	*saverm_4419/saverm_4420*	19/111
		✓	ACAAGCGGTGCG | GGAACGAATTTC	*saverm_4459 (hyaS)*	130
✓			CAAATACCTGAG | CAATTCAATTCG	*saverm_4503*	252
	✓		AACCCACCTTCG | CTCCGGAATAAA	*saverm_4510*	833
✓		✓	AAGACACGAGAT | CACAGCTATCTT	*saverm_4577*	93
		✓	AACACACGTACG | ATGACGTGTAAT	*saverm_4650*	67
		✓	CATAGACGGGCA | AAAAGGACTTTC	*saverm_4699/saverm_4700*	138/224
		✓	AAGATCCCTTCG | GGTATGAAGGTC	*saverm_4709 (disA)*	171
		✓	ACAAGACGTCCG | ACAAGGACATCT	*saverm_4913/saverm_4914*	505/-538
		✓	TAAATTCCTGAT | TGAAGGAGGAAC	*saverm_5011*	420
		✓	AAACTACCTGCT | GAGACCGACAGT	*saverm_5048 (manB)*	34
		✓	GAGAACCTGGCG | CGAAGGTATTCG	*saverm_5267 (ppe2)*	-38
	✓	✓	CAAGAACCTGAG | CAAGGGAATGGT	*saverm_5350*	5
✓	✓	✓	ACGATGCGGGCA | TGAAGGAACAAT	*saverm_5799*	-97
	✓	✓	TAACTAGATACG | CTAACGAATATA	*saverm_5837*	35
	✓	✓	AAAATACGGGAG | GAAAGGGGTGCG	*saverm_6398 (ectA)*	130
		✓	AAGCAACCTGGG | TGAACGGAAGAT	*saverm_6419 (pitH2)*	41
		✓	TACCTAACGACT | CGAACGGATTTT	*saverm_6573*	372
		✓	AAACCCGCTGCG | GGAAGGATTAGT	*saverm_6669 (oppA9)*	177
✓			GCGATATGTTCG | GTCACGAATATT	*saverm_6709/saverm_6710*	155/65
✓			AAAATCCGCACG | TCAAGGGCTTCT	*saverm_6796*	52
	✓	✓	AAGATTCCCGCA | GAACGAAAAATT	*saverm_6957/saverm_6958 (cvnA11)*	169/227
✓			CAGAGACCTGCC | CGAACTCATTTT	*saverm_7286*	-464
✓	✓	✓	AGCAGACCTTCG | CGAAGGTAGTAC	*saverm_7326*	-113
	✓	✓	ATCATACCTACT | GGTCGGTATGCT	*Unannotated ORF*	180
		✓	AGGCCATGGCCT | CCAACGAATATT	*saverm_RS39490*	-625
	✓	✓	CAGATACCTTTA | AAAACAAAGGCT	*saverm_1p77*	13

5.3 Modelling MmfR:DNA interactions *in silico*

5.3.1 Predicting MmfR residues involved in DNA interactions

To characterise how MmfR and other AHFCA receptors recognise and bind target AREs, the binding of MmfR to DNA was investigated *in silico* using the crystal structures of MmfR, in comparison with the crystal structure for the CprB:DNA complexes[69, 152] to rationalise the findings (based on the hypothesis that ArpA-subfamily repressors share common interactions formed between the repressor and DNA using the same amino acid residue positions). To provide the most accurate docking simulations, the interacting amino acid residues with which MmfR interacts with DNA were predicted. The amino acid sequence of the MmfR DNA-binding domain (DBD) was aligned against the DBDs of other TFRs using MUSCLE[233] to determine the indexing of the residue positions for which scores would be calculated (Figure 8-14, Appendix D). Residues interaction with DNA through active (electrostatic interactions) or passive (within 5 Å proximity; Van der Waals) interactions were assessed in PyMOL using the crystal structures of these aligned TFRs co-bound with DNA. From these residues a simple scoring function was generated to predict what residue positions were most likely to be involved in interactions with DNA for not just MmfR, but all TFRs:

$$Score_i = \sum \big((a + b + 2c) \times (1 + n)\big)_i$$

Values of *a*, *b*, *c* and *n* are Boolean values of 0 or 1; if the following statement is true, then a/b/c/n = 1, else 0.
a = Residue is within 5 Å proximity of the DNA molecule.
b = Residue interacts *via* backbone amide with the DNA backbone **and** does not interact *via* side-chain.
c = Residue interacts *via* side-chain with DNA backbone.
n = Residue interacts with DNA base ring.
i = Index position in MUSCLE alignment.

A significance cut-off was set at the 75th percentile of the dataset. MmfR residues which scored above this cut off (Figure 5-7) were **T30** in helix α1 (stabilisation helix), A50 and **S51** in the interhelical α1-α2 (spacer helix) loop, **V52** and **K53** in helix α2, M61 and **T62** in the interhelical α2-α3 (recognition helix) loop, **K63**, **G64**, A65, **Y67**, F68 and **H69** in helix α3, S72 in the interhelical α3-α4 (transmission helix) loop, and **K73** in helix α4 (nucleotides likely to form electrostatic interactions are in bold). Significant predictions contribute a surface area of 2,987 Å^2, and non-significant predictions 2,730 Å^2, for a total interacting surface of 5,717 Å^2. All residues in MmfR that scored above the 75th percentile were also found in the CprB crystal structures to be within 5 Å of DNA. E25, R26, S27 and I28 in MmfR helix α1 were not above the significance threshold but were within 5 Å of DNA in the CprB crystal structure, and so may also be involved in interactions with DNA.

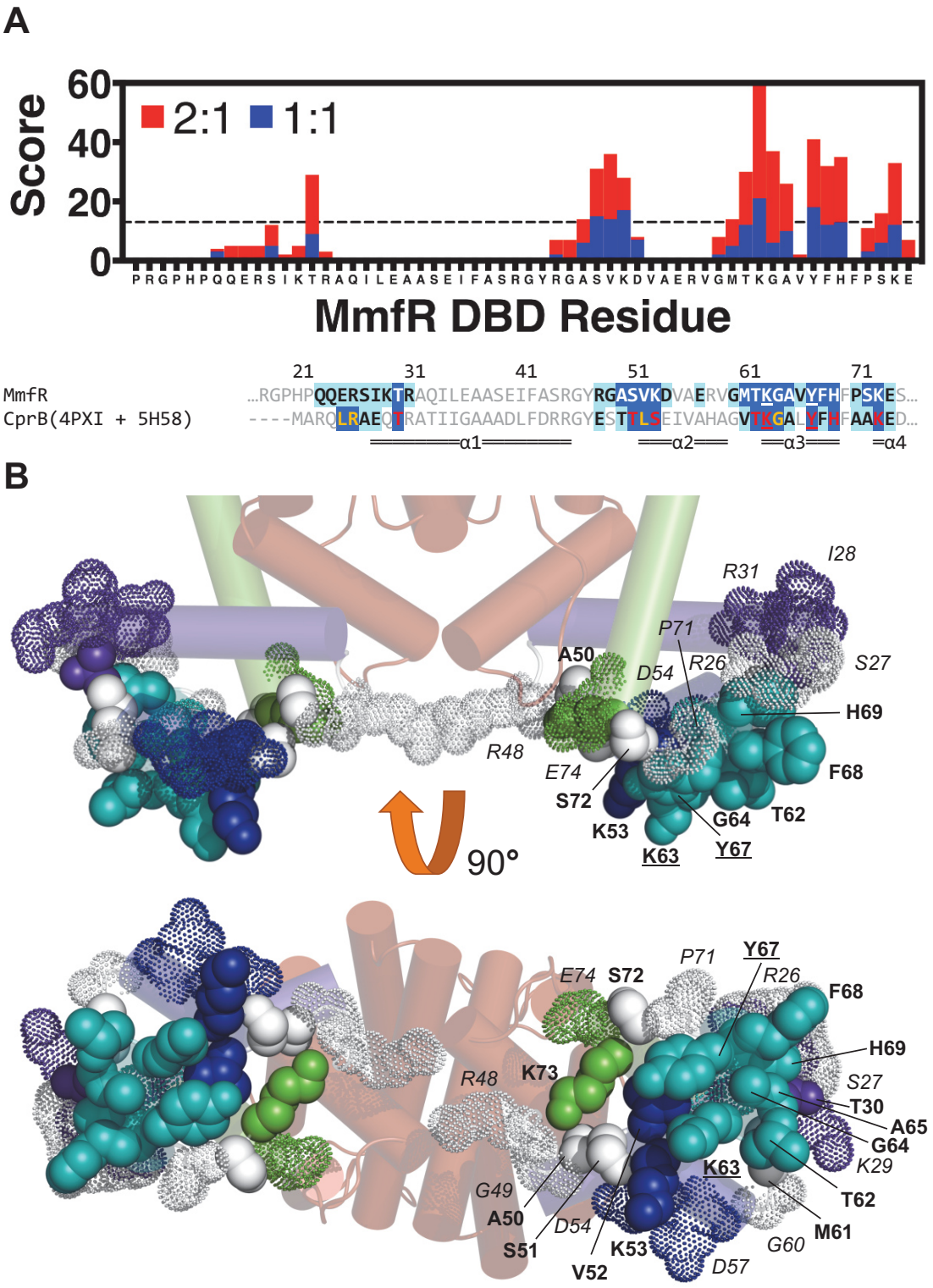


Figure 5-7 –Prediction of MmfR DNA-binding residues in the DNA-binding domain (DBD). Positional scores for each residue position in MmfR, which are annotated and shown below in alignment with CprB. Indicated by type are the scores for the binding stoichiometry of TFR dimers to DNA, generated from MUSCLE alignments shown in Figure 8-14, Appendix D**.** Dotted horizontal line at y=13 is the 75th quartile of the dataset. For MmfR **blue** indicates predictions above the 75th quartile, and cyan residues below the cut-off. For the CprB sequence, **blue** highlighted residues which interact with DNA *via* electrostatic interactions, and cyan residues which are within 5 Å of DNA. **red** residues interact *via* residue sidechains, while **yellow** for interactions *via* the residue backbone. Residues underlined interact directly with nucleotide bases. PDB identifiers for CprB are indicated in parentheses. (**B**) Structure of MmfR with strong prediction residues indicated by solid spheres and weak predictions by dotted spheres. Putative residues interacting with bases are underlined. **Purple** is stabilisation helix α1, **blue** is spacer helix α2, **teal** is recognition helix α3, **green** is transmission helix α4, **red** is the LBD, and **white** are interhelical loops.

5.3.2 Modelling MmfR residue-nucleotide interaction symmetry

ArpA-subfamily repressors are hypothesised to share a common DNA-binding mechanism with consistent stoichiometry and symmetry of dimers around the dyad centres of target AREs. While the amino acid residues which interact with DNA have been predicted for MmfR binding, determining which residues interact with which nucleotides is more difficult. For the two DNA sequences co-crystallised with CprB, four direct interactions with four different bases are formed; three when CprB is bound to the ArpA consensus sequence (*ArpA*$_{CS}$) and one when CprB is bound to the *cprB* operator (*OPB*). None of these interactions are formed in both AREs, indicating that the same residue-nucleotide interactions are not necessary for all AREs bound by CprB, and therefore exhibit some plasticity.

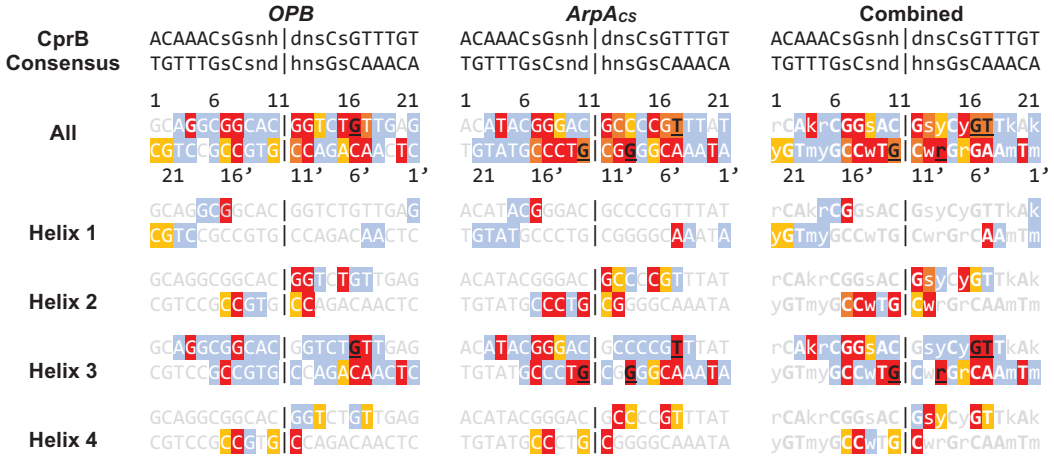


Figure 5-8 – CprB residues interacting with DNA in CprB:*ArpA*$_{CS}$ and CprB:*OPB* crystal structures. Active and passive interactions were defined for all DBD residues within 5 Å of DNA for CprB in complex with *ArpA*$_{CS}$ (PDB:4PXI) and *OPB* (PDB:5H58). Passive nucleotides are in **blue**, **red** indicates electrostatics *via* residue side chains, **yellow** *via* residue backbones, and **orange** *via* both side chains and backbones. Direct interactions with nucleotide bases are underlined.

Figure 5-8 shows a simplified map of residue-nucleotide interactions in the CprB:*ArpA*$_{CS}$ and CprB:*OPB* complexes. The C6/G17 nucleotides (±6) conserved in many ArpA-subfamily repressor ARE motifs only form one direct electrostatic interaction in the CprB:*OPB* complex *via* the nucleotide base of G17 with K43. When bound to *ArpA*$_{CS}$ interactions with this nucleotide base are lost and instead an interaction is formed with the base of the adjacent T18, but *via* Y47 instead. The other interactions in this complex are observed on the complementary strand around the dyad centre region with G'9 (K43) and G'12 (K43/Y47) (on the negative-sense strand from 5' to 3'). However, CprB is unique amongst TFRs currently in the literature in that it binds asymmetrically around the ARE dyad centre; the *ArpA*$_{CS}$ ARE has a dyad symmetry of 5'-AcAtA**C**GGGaC|GcCCC**G**TtTaT-3', but the centre of binding, equidistant between the interactions with G'9 and G'12, indicates the symmetry of dimers binding as 5'-acATAc**g**gGacG|CccCg**t**tTATn-3', one base-pair offset from the sequence dyad centre. The

T/A-rich region is still palindromic around the binding centre, but the C6/G17 nucleotides are not in alignment with ArpA-subfamily repressor recognition sites, and the same relative position is now occupied by G7/T18. Similarly, *OPB* has a sequence dyad centre around 5'-nAggcagg**C**gGcaC|GgtCt**G**ttgagtTc-3', but instead the interactions are symmetrical around 5'-aGgcaggcggcacg|gtctgttgagttCn-3'. The structures of CprB bound to both AREs have an RMSD of 0.75 Å over 715 atoms with each other; given the dissimilarity of the sequences it is probable that this offset is consistent across all sites bound by CprB. However, it is not possible to indicate if binding of CprB to DNA is representative of the ArpA-subfamily, or an exception to it.

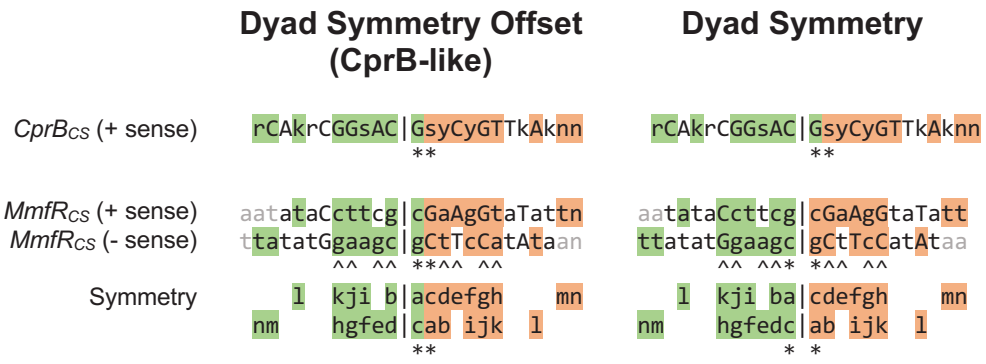


Figure 5-9 – Two symmetry models MmfR binding to a MmfR-type palindromic consensus ARE sequence. Symmetry of electrostatic interactions formed by MmfR binding to DNA is indicated by 180° rotation of symmetrical regions (green/orange) around the binding centre (indicated with asterisks). Also indicated are the ARE dyad centre (|) and probable specific interactions with specific nucleotide base-pairs (^). Bases within 5 Å of the MmfR protein are indicated in plain black text. Upper case nucleotides in the MmfR consensus ARE sequence are conserved in the nonpalindromic MmfR-type Are motif. Note that the symmetry of nucleotide interactions is not strictly as shown in the figure but simplified for demonstration of concept. Symmetry for some interactions are not observed in CprB crystal structures due to non-100 % palindromacy of the sequence; therefore, some interactions are only implied by symmetry.

Therefore, for MmfR binding to DNA there are two hypothetical configurations of binding to DNA (Figure 5-9); either the binding of two MmfR dimers is symmetrical with respect to the ARE dyad centre, or the two MmfR dimers binds with a symmetry offset from the dyad centre like CprB. In addition, because ARE sequences are palindromic, offset may be in either direction from the dyad centre. Assuming an equal distribution of electrostatic interactions around the centre of MmfR binding symmetry, many of nucleotide base-specific interactions are more likely to be formed in the region around the dyad centre, rather than at the distal ends of the ARE sequences, consistent with the localisation of helix α3. This includes the bases in the regions around the C7/G18 (although only in a half-dyad for the CprB-like model), and in the T/A rich regions 3-bp to 4-bp from the dyad centre, which are proposed to be significant for AHFCA receptor recognition. By contrast, fewer interactions would be formed between MmfR and the T/A-rich distal regions, with interactions here presumably to stabilise the interaction of two homodimers binding to the ARE via the stabilisation helices. The

prevalence of this distal A/T-rich region in ArpA-subfamily repressor ARE motifs would initially indicate that these nucleotides have some functional significance for recognition, which, if not by forming direct interactions with the repressor, may instead be reflected in the bendability of the target sequence to enable the stable repressor:DNA conformation to be adopted.

5.3.3 Docking of MmfR to the MmfR-type consensus sequence

The crystal structure of MmfR in the absence of any ligands was docked in HADDOCK to the 24-bp MmfR consensus sequence 5'-AATATACCTTCGCGAAGGTATATT-3', with the flexible DNA model structure generated in 3D-DART. The residues interacting either actively or passively where specified based on the previous predictions, but specific nucleotide interactions were not defined in the simulation parameters to replicate *in situ* the localisation and docking of MmfR helix α3 into the same major groove regions of DNA as observed in the CprB crystal structures. The docking simulated generated 145 structures, grouped into 17 structure clusters. In the best scoring structure (Figure 5-10) the range of nucleotides within 5 Å of the MmfR dimer was 21-bp (of which the range of electrostatic interactions was 17-bp), and the surface area of residues within 5 Å of DNA was 2334 $Å^2$. MmfR localised to the same position of DNA as CprB Dimer A (chains A and B) relative to DNA, with an RMSD of 23.94 Å for 115 atoms when aligned. However, only the chain A α3 helix was docked into the DNA major groove, while the α3 helix of chain B was docked over the minor groove distally adjacent from the expected target major groove. Alignment with the CprB structure bound to DNA shows distortions of the phosphate backbone resulting in an expansion of both major grooves in the CprB structure that is not accounted for in the simulation.

All residues that were predicted to form electrostatic interactions with DNA did so except for E25, R26, T30 and H69. Interactions that weren't observed in CprB were formed by A65 and S72 in MmfR chain B, and R48 in both MmfR chains. MmfR Chain A R26 was the only α1 helix residue within 5 Å of DNA; twisting of the dimer interface, and the resulting distortion of DNA may bring R26 and T30 into closer proximity to DNA, but the distortion must be larger than that for CprB based on the relative angle of the stabilisation helices. Any interactions between the MmfR stabilisation helix and DNA likely only occur after both α3 helices are properly docked into the major grooves. Similarly, the H69 residue of MmfR is within 5 Å proximity of DNA, but due to the poor docking of chain B to DNA, the expected electrostatic interaction in the second major groove is not observed (in CprB this histidine residues interacts through chain B only, and not through chain A).

In MmfR chain A, the R48 sidechain found in the interhelical α2-α3 loop forms four electrostatic interactions with dyad centre of DNA; three with G'12 and one with C'13 (5'-n_{10}-C**GC**G-n_{10}-3'). Meanwhile, chain B residue R48 forms two interactions with G'12. It is anticipated that the R48 residues of the second MmfR dimer binding would also interact with the dyad centre residues symmetrically. Because the middle point of the two R48 residues forms a dimeric pseudo-centre, another pseudo-centre between the pairs of dimers may suggest the centre of symmetry for electrostatic interactions. However, this pseudo-centre does not line up with the centre of symmetry between the CprB dimers. If this R49 interaction is significant, it could mean a difference in angle between dimers is observed for MmfR through a ~1-bp rotation of dimer B around the major groove. Alternatively, this interaction may not be significantly specific for the dyad centre, due to the flexibility of both the loop and the arginine side-chain. Finally, all nucleotides fully conserved in the MmfR nonpalindromic ARE motif form electrostatic interactions with MmfR either directly or *via* their palindromic position, except for the G14 residue which only forms interactions *via* symmetry around the asymmetric model. MmfR chain A Y67 forms a direct interaction with the G18 base (C7 dyad opposite; the highly conserved C/G nucleotide in ArpA recognition motifs), which is also formed in the same position in CprB. In the symmetric model, C7/G18 are involved in anchoring around the dyad centre.

That MmfR docks *in silico* to the same region as CprB *in vitro* may suggest that other ArpA-subfamily repressors have a similar distribution of dimers around bound AREs. That only one α3 helix in the MmfR dimer initially fits into the major groove indicates that a conformational rearrangement of the dimeric interface is required to fit the second α3 helix into the distal major groove, as described in the clamp-and-click model of binding. It is still unclear how the binding of two dimers occurs, with speculative models depending on an interpretation of the model proposed in Chapter 4.4.4. The α3 helices of both dimers may fit into the distal ARE regions simultaneously, followed presumably by interdependent intradimeric twisting of each dimer to under twist the DNA and allow major groove docking of the second α3 helix. In the alternative sequential model binding of the first dimer may occur as described above, which exposes the distal major groove and allows the second dimer to now recognise the target ARE through fitting of one of the α3 helices into the major groove where it couldn't before, possibly due to poor A/T composition which prevents DNA undertwisting. An actual structure of MmfR in complex with DNA would be able to expand upon this hypothesis further.

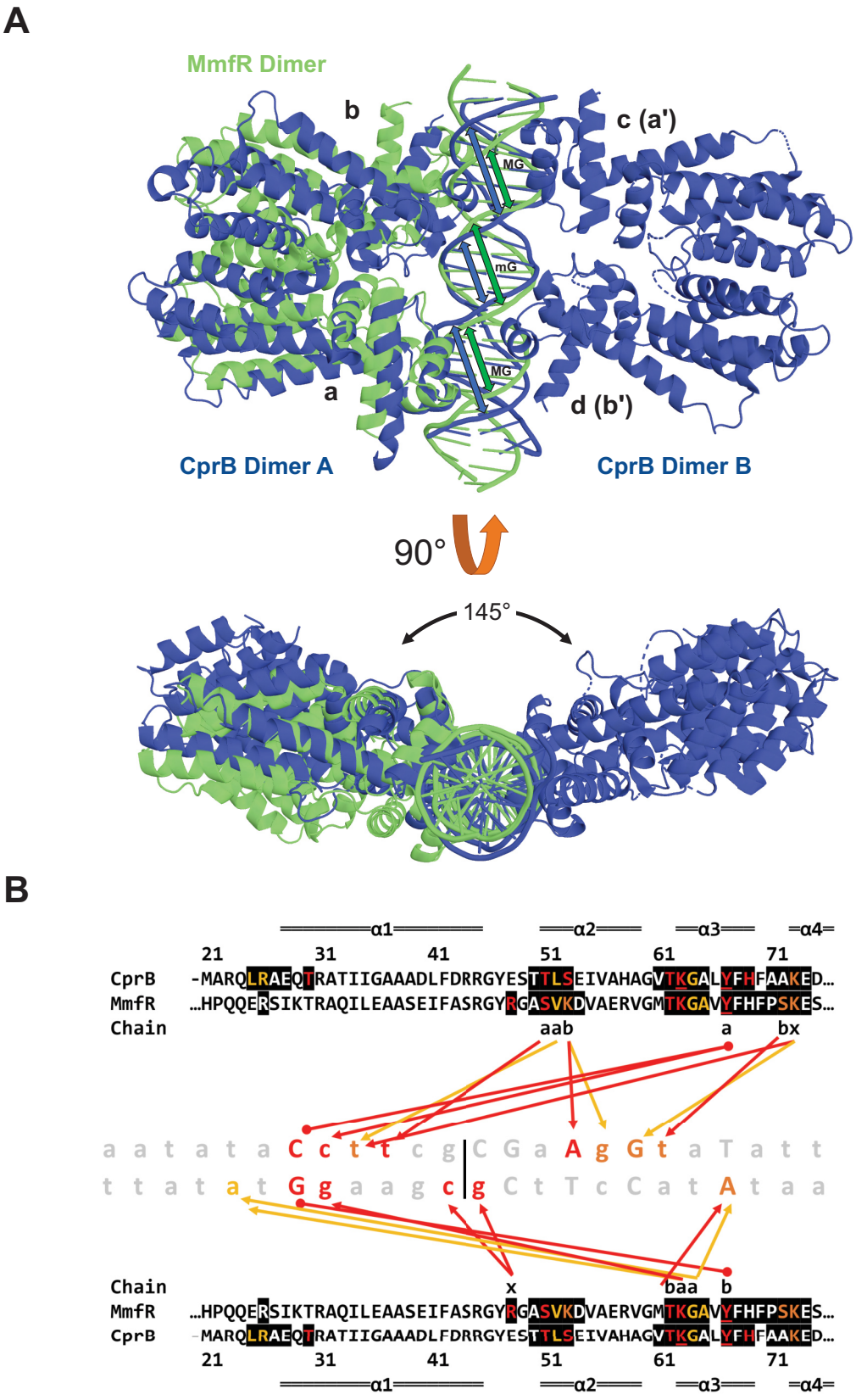


Figure 5-10 – HADDOCK simulation of MmfR binding to the MmfR-type consensus sequence. (**A**) MmfR (**green**) was docked and then aligned with the crystal structure of CprB bound to the *ArpA*$_{CS}$ (PDB:4PXI; **blue**). CprB monomers are indicated in lower case letters, with a' and b' indicating symmetry. Arrows indicate relative distances of major grooves (MG) and minor grooves (mG) for each structure. (**B**) Residue-nucleotide interactions map. For sequences/arrows **red** indicates residue side-chain interactions, **yellow** indicates residue-backbone interactions and **orange** indicates both. Circle-tipped arrows indicate direct interactions between MmfR residues (underlined) and nucleotide bases. Chains of the MmfR residue interacting with DNA are indicated for Chain A (a), Chain B (b), and both chains (x).

Possible electrostatic interaction symmetries (Figure 5-11) are also hypothetical at this stage, but a further assessment of symmetry should be considered when a crystal structure of MmfR:DNA is made available. Docking of a second MmfR dimer was attempted, but a docking error in the annealing of the second dimer to DNA resulted in a failed simulation. Because the second dimer couldn't be docked, and because some expected interactions were missing due to the chain B α3 helix not docking into the second major groove as expected, it is difficult to assign a symmetry model for MmfR binding to the ARE sequence. An additional structure model was generated by duplication and alignment around the complementary ARE strand (Figure 8-15, Appendix D), but as can be expected it was only in acceptance with the centred symmetry model, and the second dimer was not offset from the dyad centre as has been observed in all other crystal structures of TFRs binding as pair of dimers. The prediction of which amino acids interact with DNA is likely more reliable than the nucleotides they are predicted to interact with, given the different number of interactions that are formed between CprB and the two sequences it binds to. It is highly probable that the positions of these interactions change depending on the composition of each sequence.

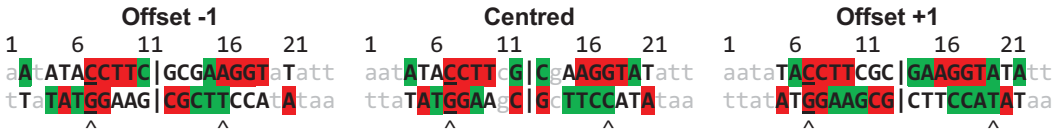


Figure 5-11 – Possible symmetry model for the binding of MmfR to the palindromic MmfR-type consensus motif. Shown are each model for (**A**) native MmfR and (**B**) MmfR generated from chain A and B of DNA-bound CprB (PDB:4PXI) using Phyre[2]. Symmetry of electrostatic interactions formed by MmfR binding to DNA is indicated by 180° rotation of simplified symmetrical half-sequences around the binding centre (|), where **red** indicates the nucleotides interacted with in the docking simulation and **green** the symmetrical interactions.

5.3.4 Comparing ArpA-subfamily repressor DNA-binding domains

Phylogenetic analysis was performed to group the DBDs of ArpA-subfamily repressors described in Chapter 5.2, as well as the DBDs of other TFRs with have been co-crystallised with DNA (Figure 5-12). Also included are the previously characterised SbbR (41 % sequence identity with MmfR), the *Streptomyces tacrolimicus* tacrolimus biosynthesis regulator Gbr (40 % sequence identity with MmfR)[234], and an uncharacterised homologue SSHG_01258 in *S. albus* (40 % sequence identity with MmfR), which has been proposed to bind to the MmfR-type AREs during *in vivo* studies performed by Dr Styles[149]. As expected, ArpA-subfamily repressor DBDs clustered together in the phylogenetic tree by receptor roles, with GBL/Pseudo-GBL and AHFCA/AHFCA-like/Pseudo-AHFCA receptors diverging into two separate clusters. CprA and CprB did not cluster together with the other GBL receptors but instead with the Pseudo-GBL receptors, albeit in a separate clade. This arrangement is debatable, given that the bootstrap percentages for this arrangement is quite low.

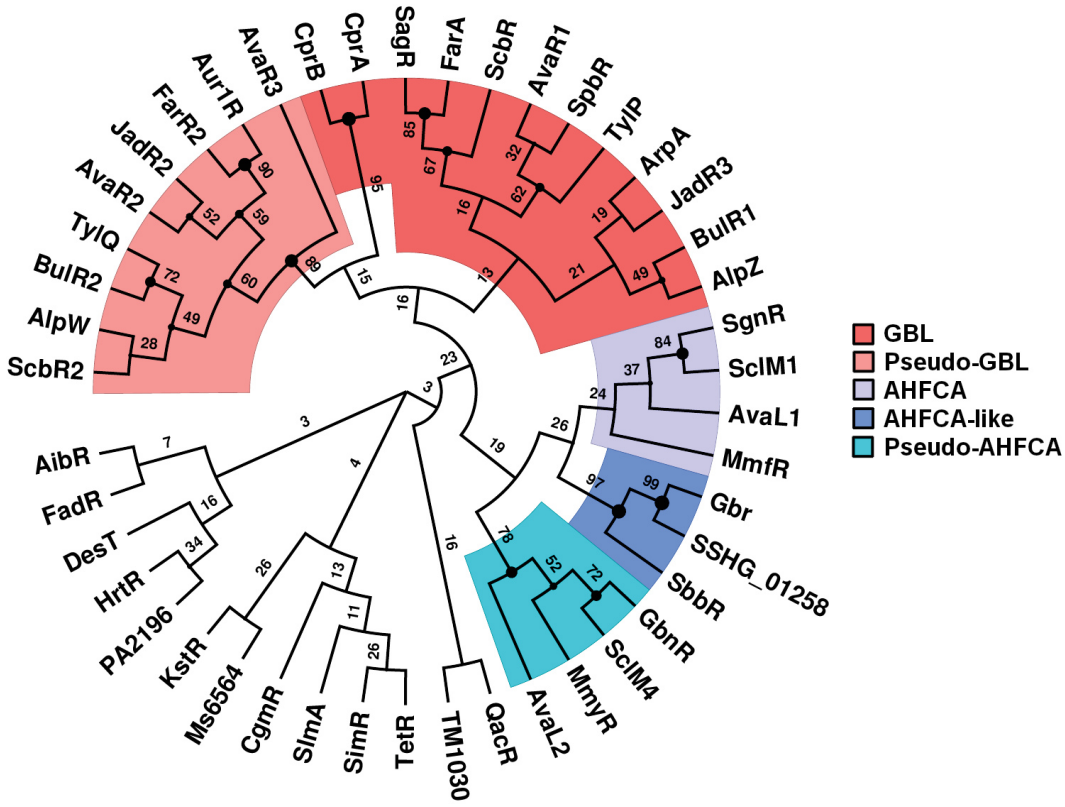


Figure 5-12 – Unrooted Maximum Likelihood bootstrap consensus tree of ArpA-subfamily repressor DNA-binding domains (DBDs) inferred over 1000 iterations. Sequences of DBDs where aligned using MUSCLE, and the tree generated using MEGA X. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT matrix-based model, and then selecting the topology with superior log likelihood value. ML Heuristic method used was Subtree-Pruning-Regrafting level 5. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the branches.

Figure 5-13 shows the alignment of the DBDs of these ArpA-subfamily repressors. Residues interacting electrostatically with DNA are conserved across the ArpA-subfamily repressors in positions aligned with MmfR R26 (except for AlpZ), T30 (except for AlpZ), G64, H69 (except for TylQ) and K73. The residue position at L52 is always occupied by a hydrophobic residue. The most disparately conserved region between ArpA-subfamily repressors is between helix α1 and the N-terminal region of helix α2, even between ArpA-subfamily repressors of the same receptor group. Helix α3 of the AHFCA receptors is conserved with the GBL receptors except for a single L66V substitution. The pseudo-AHFCA receptor α3 helix is conserved with the AHFCA receptor α3 helix with the exception of F68G and L66 which is conserved with all other receptor groups (except for the AHFCA-like receptors).

In helix α3 T62 is conserved in all ArpA-subfamily repressors except for the pseudo-GBL receptors, where it is substituted with serine. K63 which forms interactions directly with nucleotide bases is again conserved in all receptor groups except for the pseudo-GBL and AHFCA-like receptors, where it is replaced with either asparagine, proline or serine in the former and serine in the latter. M61, despite not being directly involved in electrostatic

interactions, is likely to be involved in directing specificity to the target DNA sequence; mutation of this position, which is a conserved valine in the GBL and pseudo-GBL receptors shown in the figure, abolishes DNA-binding activity in ArpA to the *adpA* site when substituted with alanine[235]. This is methionine is present in all AHFCA receptors and MmyR but substituted and with hydrophobic residues in all other receptor groups except for the AHFCA-like receptors, where it is substituted with a charged amino acid. The other residue which forms interactions with nucleotide bases, Y67, is conserved in all receptor groups except for the pseudo-GBL receptors, where it is substituted with histidine.

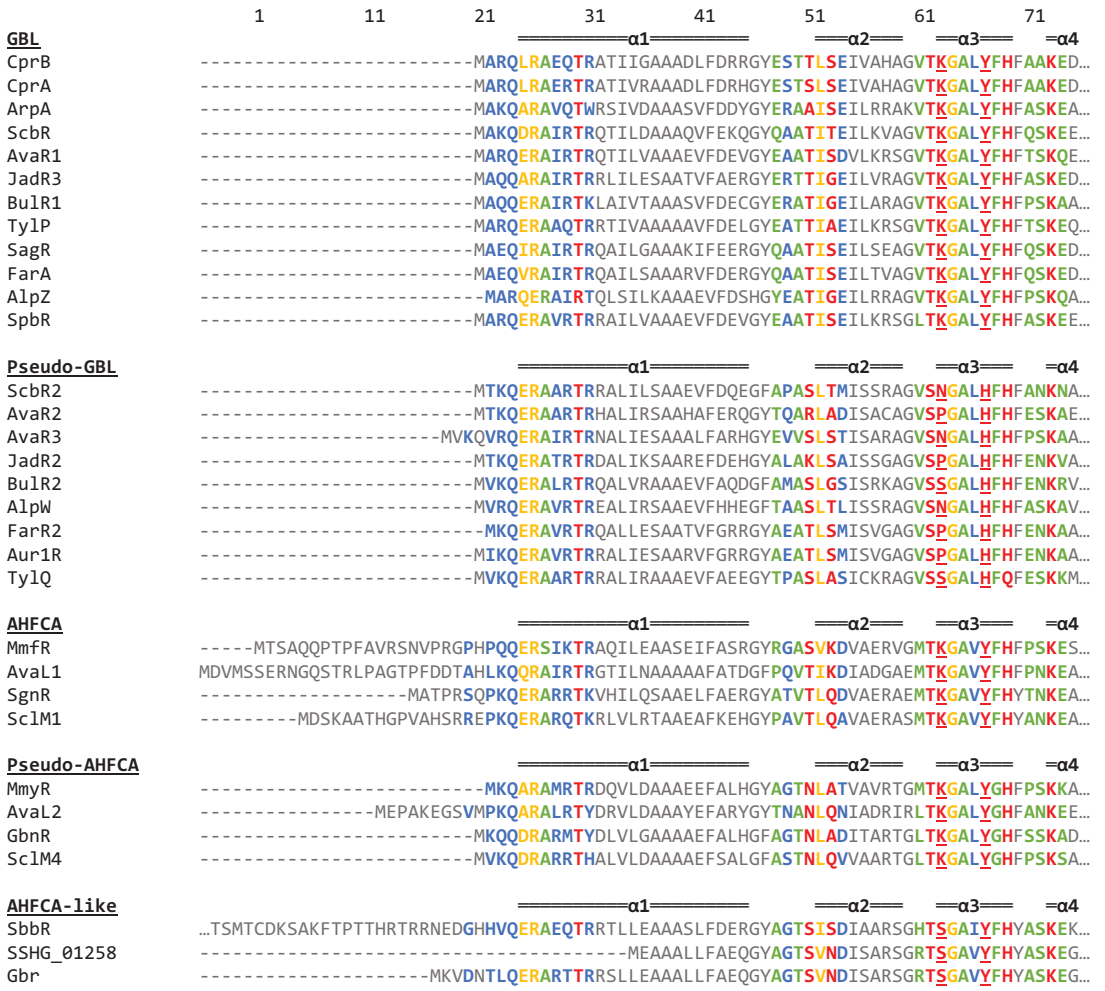


Figure 5-13 – MUSCLE alignment for ArpA-subfamily repressors indicating residues most likely involved in electrostatic interactions with DNA. Residues interacting *via* sidechains are **red**, residue interactions *via* backbone are **yellow**, passive residues found in CprB in **green**, and passive residues based on other TFRs in **blue**. Residues which can potentially interact with nucleotide bases are underlined. Amino acid position is based on the sequence of MmfR. Position of DBD helices based on MmfR structure are indicated with =.

The passive residues as well non-surface residues which affect the surface geometry may also have some measurable effect on the specificity of ArpA proteins, but these contributions are difficult to predict. R48 in MmfR is not conserved in other MmfR homologues, where it is substituted with proline in AvaL1 and SclM1 and alanine in SgnR. It is not known what affect

this arginine residue has on DNA affinity: it is possible that mutagenesis of MmfR homologues to contain this arginine may improve affinity with DNA. Similarly, P71 and S72 are also not conserved, but substitutions at this position are not likely to be significant.

5.3.5 Consideration of other factors influencing DNA recognition

Differences in α3 helix residue composition may affect how well the helix docks within the major groove of DNA, leading to variations seen in the ArpA recognition motifs, but it does not determine absolute recognition. ScbR, which has an identical recognition helix to CprB, does not bind the *OPB* operator as both CprB and ScbR2 do, but does bind the *cprA* operator, and CprB can bind some of the ScbR target sequences. SUPER alignment of TFR crystal structures show that despite sequence homology of CprB to MmfR, CprB is only the 4th most structurally similar TFR to MmfR (4.3 Å RMSD), after FadR (2.6 Å RMSD), TylP (3.6 Å RMSD) and SlmA (4.1 Å RMSD). The physical arrangement of helices in the MmfR DBD is more closely aligned with FadR and TylP than CprB; helix α1 is almost flat in MmfR, FadR and the other TFRs, but in CprB the stabilisation helix is tilted downwards at an ~18° angle, and with it the spacer and recognition helices are subtly reoriented (Figure 5-14). There was no significant correlation between the relative angles of helices α4 and α1. Whether a difference in the tilt of helices within the DBD has some effect on DNA recognition is not clear; as the DBD is the most flexible domain of the protein, there is likely some range of motions through which the helices can rotate relative to the fixed position of helix α4 (through the formation of the hydrophobic core of the DBD).

Surface distribution of electrostatic potential across the surface geometry of the DBD likely contributes to DNA recognition also. Despite differences in helix angles, the surface distribution of electrostatic potential over the DBD of MmfR and CprB are more similar than other TFRs when compared with FadR, SlmA and TetR (Figure 5-15). TylP is more deviated in its distribution of electrostatic potential, in part because of the C-terminal extension spanning to the DBD, which differentiates its binding mechanism from the other ArpA-subfamily repressors, but at the same time shares some characteristics; positive electrostatic potential runs parallel across the length of the DBD, whereas in FadR, SlmA and TetR the positive electrostatic distribution is either interrupted or antiparallel to the length of the DBD. Other ArpA-subfamily repressors are likely to have positive charge distributed parallel down the length of the DBD, but whether the distribution is contiguous like MmfR or less well defined as in TylP remains to be determined. Phyre^2 modelling of structures however could serve as good predictions of electrostatic distribution.

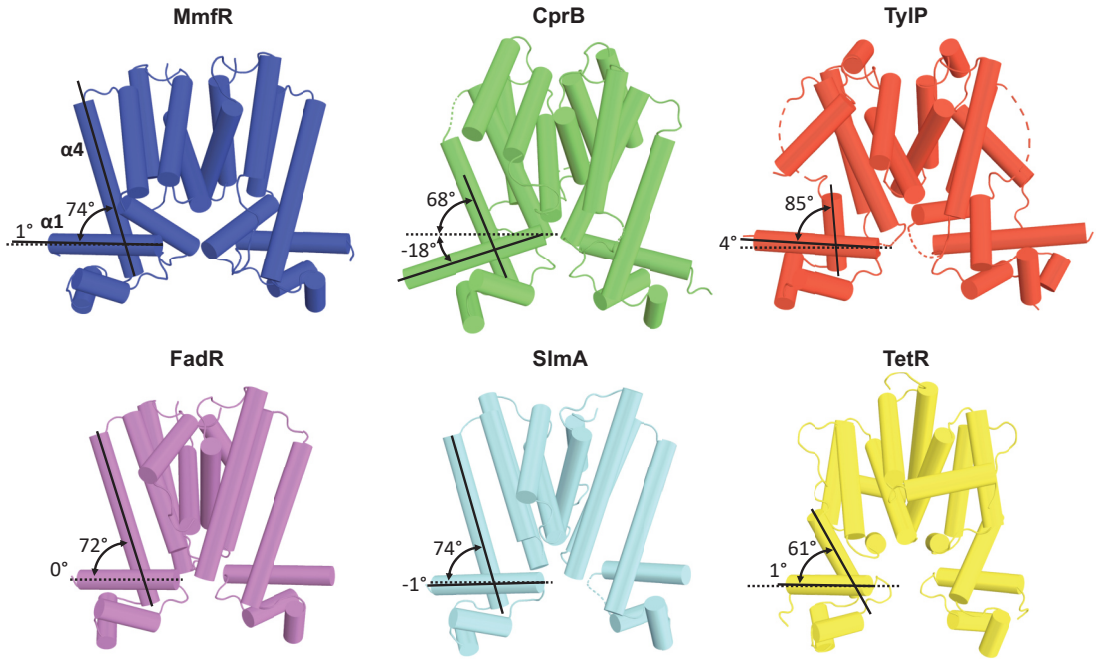


Figure 5-14 – Tilting of helices α1 and α4 in TetR-family transcriptional repressors. The top row are ArpA-subfamily transcriptional repressors, while the bottom shows non-ArpA-subfamily repressors, including the structurally similar repressors *Bacillus halodurans* FadR and *E. coli* SlmA, as well as the family representative repressor *E. coli* TetR. A flat horizontal line was rooted to the C-terminus of helix α1, and the angles for helices α1 and α4 (solid lines) to the line were measured. PDB identifiers can be found in Table 1-2, Chapter 1.

Another important factor is distal effects of residues in the LBD on the dimeric interface; TFRs will have different energetic requirements for twisting around their dimer interface during DNA-binding that is dependent on the composition of the dimeric interface, which in turns drives specificity for sequence with different "bendability" profiles. It has been shown previously that *S. griseus* strain H01 encodes an ArpA$_{P115S}$ mutant that does not bind DNA, but still binds A-factor[236]. This proline residue, located in α6-α7 interhelical loop, is conserved in both CprA/CprB and in the AHFCA and pseudo-AHFCA receptors. A mutation to serine may result in new interactions between the hydroxyl group of serine and helix α4, which could interfere with subsequent twisting of the dimer interface. Because the proline residue lies in a flexible loop, this residue position in CprB and MmfR (P121 and P141 respectively) differs in their crystal structures. While the pyrrolidine ring in CprB is oriented towards helix α7 and forms no electrostatic interactions, in MmfR the ring is oriented towards helices α4 and α6, allowing the carbonyl group to form an electrostatic interaction with the sidechain of LBP residue Q130, which in turn interacts with Y85 to stabilise binding of AHFCAs. Mutation of this residue in MmfR, and other residues in the DBD, may affect the DNA-binding ability of the MmfR repressor.

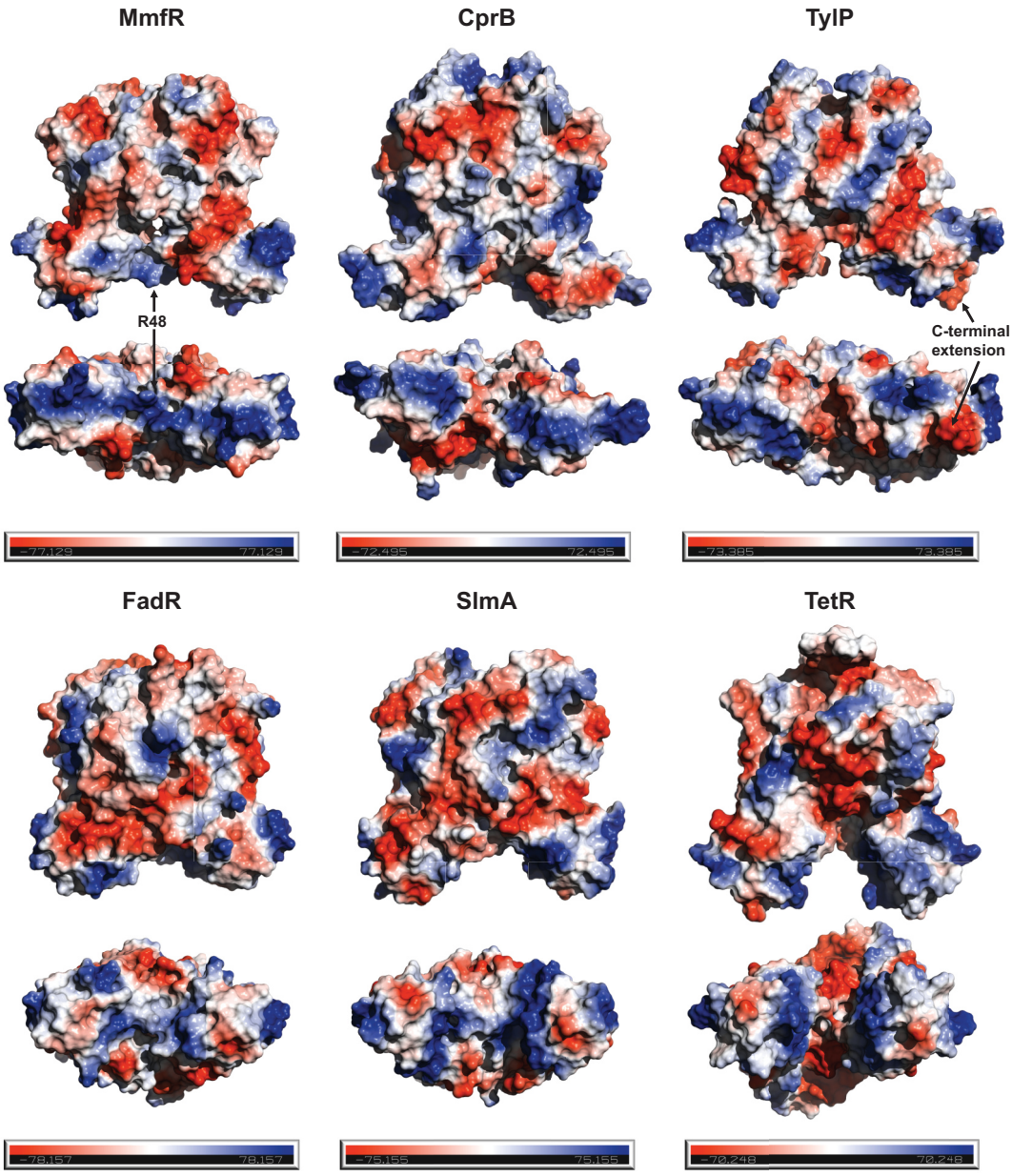


Figure 5-15 – Surface electrostatic potential of the DBD for select TFRs available on the PDB. The top row are ArpA-subfamily transcriptional repressors, while the bottom shows non-ArpA-subfamily repressors, including the structurally similar repressors *Bacillus halodurans* FadR and *E. coli* SlmA, as well as the family representative repressor *E. coli* TetR. Positive potential is indicated in **blue**, and negative potential in **red**. Shown are the front-facing orientation of each TFR, and below this and oriented view of the DBD through a -90° rotation of the X-axis. PDB identifiers can be found in Table 1-2, Chapter 1.

5.4 Discussion and conclusions

A MmfR-type ARE motif universal for the four characterised MmfR homologues in *S. coelicolor* A3(2), *S. avermitilis*, *S. venezuelae* and *S. sclerotialus* was determined in both palindromic and nonpalindromic configurations, which in its palindromic configuration has a consensus sequence of 5'-AAnATACCTTCG|CGAAGGTATnTT-3'. This ARE motif shared several features with other ArpA-subfamily repressor binding sites, including a conserved degeneracy of the dyad centre (which is less degenerate in the MmfR-type ARE motif), a highly conserved C/G nucleotide 6-bp from the dyad centre and A/T rich distal regions with high palindromic frequency. All ARE motifs have consensus of A/T (W) and C/G (S) dinucleotide blocks, but these differ between the MmfR-type ARE, which has a consensus of 5'-WWWS**W**S|S**W**SWWW-3' (TT in the -3-bp to -4-bp region and AA in the +3-bp to 4-bp region), and all other ArpA-subfamily repressor AREs, which have a consensus of 5'-WWWS**S**S|S**S**SWWW-3'. Future experiments could examine this interpretation of nucleotide arrangements by assaying MmfR binding to libraries of consensus sequences with dinucleotide block positions swapped. The only other ArpA-subfamily repressor for which the binding motif is significantly different from other ArpA-subfamily repressors is SbbR, an AHFCA-like receptor which share even less similarity in its ARE recognition motif with other ArpA-subfamily ARE motifs than the MmfR-type ARE motif.

Transcriptomics and other chip-based methodologies are often used in combination with molecular biochemical assays to identify possible regulatory sites in deleterious or overproducing mutant strains. However, some sites cannot be identified in this manner; a clear example being the Δ*mmfR* deleterious mutant of *S. coelicolor* J1056 (SCP1$^+$, SCP2$^-$*hisA1 uraA str1 pgl-1*), which sees no phenotypic change from the wild-type in methylenomycin production despite having confirmed regulatory sites within the *mmy* gene cluster[67]. It is hypothesised that this lack of phenotypic change occurs because of redundant repression of the *mmf*/*mmy* biosynthetic genes by another repressor, potentially either the pseudo-AHFCA receptor or another ArpA-subfamily repressor. In the *cpk* gene cluster of *S. coelicolor* A3(2) deletion of the GBL receptor gene *scbR* similarly results in no production GBLs as well delaying RED production[98]. It is now known that the divergent promoter of *scbA*/*scbR*, which is bound by ScbR, is also bound by the ArpA-subfamily repressors ScbR2 and CprB[69, 76]. From the bioinformatic analysis of ArpA-subfamily ARE motifs it is clear that many of these repressors have some ability to recognise the AREs of other ArpA-subfamily proteins encoded within the same genome; this redundancy being a possible reason why these targets are not easily observed in deleterious mutants. Based on homology with known ArpA-subfamily pleiotropic regulators the AHFCA/Pseudo-AHFCAs receptors may also function as

pleiotropic regulators, presumably in coordinating secondary metabolite biosynthesis with other ArpA regulators.

Using the MmfR-type ARE motif many putative AREs were predicted bioinformatically within the genomes of *S. coelicolor* A3(2) and *S. avermitilis*, with just over a third being in the upstream regions of annotated ORFs. Several AREs were predicted within the AHFCA-regulated gene clusters; in addition to the already confirmed targets of the AHFCA receptors, putative new regulatory targets include *mmyX* and *mmyQ* in the *mmy* gene cluster of *S. coelicolor* A3(2) and *saverm_2278*, *fabF*, *fabI2*, *saverm_2300* and *fadE10* in the azoxy compound gene cluster of *S. avermitilis*. The predicted MmfR-type AREs were also cross referenced with known of pleiotropic targets of ScbR/ScbR2[76], and significantly matched with previously characterised AREs bound upstream of *scbA*/*scbR*, *cpkO*, *cpkM*, *cprA* and *cprB* in *S. coelicolor* A3(2), and upstream of *aveT*, *avaR1/aco* and *avaR2* in *S. avermitilis*. One of the issues made evident during bioinformatic analysis of ARE is that many of the AREs bound ArpA-subfamily repressors outside of their encoding gene clusters are distinct from the gene cluster AREs by apparent increases in degeneracy/palindromacy. This is accounted for the presence of AREs encoding only a single ARE-like half-dyad, which are more prominently bound by the pseudo-GBL receptors. Given that the only known AREs bound by MmfR-type repressors are highly palindromic and well conserved, it is not clear to what extent MmfR-type repressor can bind these half-dyad AREs. Some efforts were made to account for these half-dyad predictions in the bioinformatic analyses, but this reflects also in an increased difficulty in the predication of half-dyad AREs, as with only a half-dyad to search for the number of nonsignificant predictions using bioinformatics would surely increase. As such, using the current analyses not all possible AREs may be accounted for.

Bioinformatics does have some advantages, one being analysis of deleterious mutants in which not all targets may be observed because of redundant transcriptional repression, by ArpA-subfamily repressors or otherwise. In addition, some *Streptomyces* species such as *S. avermitilis* and *S. sclerotialus* have poor genetic tractability, so generating mutant strains can sometimes be a difficult task. However, with the possibility of AREs not being identified in bioinformatic analysis, and without experimental data to corroborate the validity of these *ab initio* predictions, it is clear that bioinformatics comes with its own disadvantages. Thus, bioinformatics and transcriptomics should be treated as complimentary to one another, which may otherwise be compromised when performed in isolation.

Analysis of the crystal structures of CprB and other TFRs co-crystallised with DNA indicate that many of the same DBD residue positions frequently interact with DNA. It is predicted that MmfR (and other AHFCA receptors) uses the same positions of amino acids in the DBD as in CprB to form electrostatic and Van der Waals interactions with DNA. However, making one-to-one predictions based on the crystal structure of CprB in complex with DNA is difficult given that CprB that orientation of their DBD helices are dissimilar when compared to other TFRs, that CprB does not binding to target AREs with symmetry matching that of the proposed ARE dyad centre, and that CprB exhibits plasticity in the interactions it forms with DNA dependent on the sequence of the DNA molecule. Residue-nucleotide interactions formed in the *in silico* docking simulations MmfR binding to the MmfR palindromic ARE consensus sequence 5'-AATATACCTTCGCGAAGGTATATT-3' however were similar to those formed by CprB, with many non-specific interactions formed with the conserved region around the dyad centre, with specific amino acid:nucleotide base interactions formed with the conserved C7/G18 residues serving as anchors for the binding of MmfR to the sequence. The MmfR structures, without defining the position of DNA to which they bound, localised to the same regions of DNA as CprB binds, validating the strength of the prediction algorithm. The A/T-rich distal regions were not proposed to form specific interactions with the repressor; the distribution of GC and AT nucleotides within these AREs could be important for DNA recognition, as these nucleotides may influence the "bendability" of DNA, manipulated by twisting of the TFR dimer interface as required for binding.

That CprB binds as a pair of homodimers, and that MmfR docked to the same position, gives credence to the hypothesis that MmfR also binds as a dimer of dimers. In addition, docking of the ligand-free MmfR structure showed poor fitting of the second recognition helix into the major groove, whereas the CprB-modelled MmfR had reasonable fit of both helices. This may indicate that the docked structure with only one helix in the major groove is an intermediate complex which then undergoes twisting of the dimer interface to allow fitting of the recognition helix into the major groove. While it is clear that the DBD sequences of ArpA-subfamily repressors has some influence on the recognition of their target AREs, other factors, such as surface electrostatics, DBD helix geometry and dimeric interface stability likely play other roles in determining the specificity of different ArpA-subfamily repressors

6 Investigating the structure-activity relationship of AHFCAs with their target receptors

6.1 Aims and Strategies

Eleven AHFCA signalling molecules have thus far been identified *in vivo* in both *Streptomyces* spp. and *E. coli* (Figure 6-1) through either activation of target gene clusters by deletion of MmyR-type regulators or heterologous expression of AHFCA biosynthetic genes[95, 111, 112], in addition to a number of synthetic AHFCAs produced *via* organic synthesis[139]. Three receptors with AHFCA receptivity have also been confirmed; *S. coelicolor* A3(2) MmfR[139], *S. venezuelae* SgnR[148], and now *S. avermitilis* AvaL1, with gel shift assays demonstrating that the AvaL1 repressor is released from $saverm\_2301_{ARE}$ in the presence of AHFCA6 (Figure 3-10, Chapter 2.5.4). Each of these species produces different AHFCA signalling molecules, with *S. coelicolor* A3(2) producing AHFCA(1-5)[95], *S. venezuelae* producing AHFCA(5-10) (excluding AHFCA8E)[111], and *S. avermitilis* putatively producing AHFCA(3, 5-7) (determined through expression of *avaABC* AHFCA biosynthetic genes in the heterologous host *S. lividans* TK24 pESAC13A-2H1 Δ*avaL2*)[112].

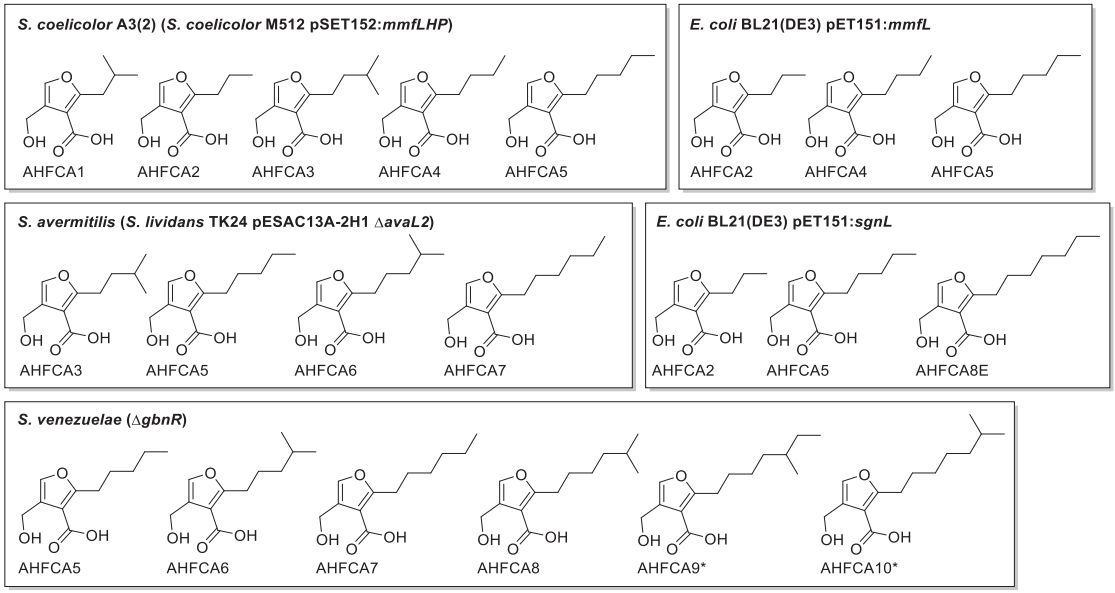


Figure 6-1 – (Repeated figure) Structures of AHFCA molecules produced *in vivo* by *Streptomyces* spp. and *E. coli* transformed with AHFCA biosynthetic genes. Figure is identical to Figure 1-8, Chapter 1. Asterisks indicate novel AHFCAs produced in feeding studies with d10-labelled leucine and isoleucine.

Much of what is known about the interaction of AHFCA effectors with their target receptors, as between AHFCA receptors and DNA, is based on either semiquantitative *in vitro* analysis *via* gel shift assays or *in vivo* quantitative analysis. Here a quantitative *in vitro* characterisation of the interaction between AHFCA inducers and their receptors are reported, using a Streptavidin (SA) chip-based SPR protocol to measure the ability of AHFCA signalling molecules to inhibit the ability of AvaL1 to bind to SA-immobilised 5'-biotinylated hairpin dsDNA oligonucleotides. In addition to AHFCA signalling molecules, two control ligands were also assayed; the *S. coelicolor* A3(2) antibiotic methylenomycin A (MmA) and the *S. coelicolor* A3(2) 6*R*-hydroxy-type GBL SCB1 (Figure 6-2). All AHFCAs ligands for these experiments used have been synthesised by Dr Zhou using the protocol described by Davis *et al*[237], and expanded upon by the work of Dr Malet[139]. Synthesis of SCB1 was also performed by Dr Zhou, using a modified route based on the synthesis of A-factor as reported by Morin *et al*[148, 238]. MmA was purified from the MmA-producing heterologous host *S. albus* J1074 strain W303 by Dr Idowu[94].

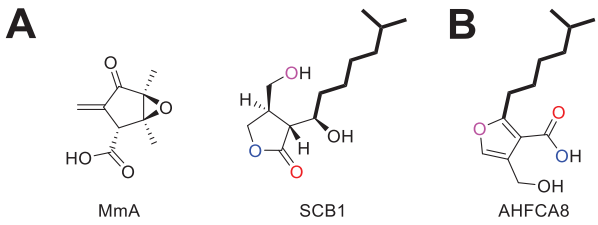


Figure 6-2 – *S. coelicolor* A3(2) effectors used as negative controls in SPR experiments. (**A**) Chemical structures of the negative controls methylenomycin A (MmA) and the 6*R*-hydroxy-type GBL SCB1. (**B**) Structure of AHFCA8 produced by *S. venezuelae*, with which SCB1 shares a common biosynthetic pathway. Alkyl chain is indicated in bold, with analogous heteroatoms indicated by colour.

The structure-activity relationship (SAR) of AHFCAs was investigated using this SPR data, in combination with previous experimental data collected by past Corre members and *in silico* analyses of AHFCA effectors binding to their target receptors, based on the crystal structure of MmfR in complex with AHFCA2. To this end, Phyre^2 was used to build homology models of AvaL1, SgnR and *S. sclerotialus* SclM1, and docking analyses were performed using AutoDock VINA (ADV). Using this data, comparisons of AHFCA SARs between MmfR and AvaL1 were made. *In silico* structure-led ligand design of AHFCA analogues for activity against MmfR using the methods developed herein were also considered.

6.2 Inhibition of DNA-binding by effectors using SPR

6.2.1 Pilot Experiment: Inhibition of AvaL1 DNA-binding by AHFCA4

A pilot experiment was performed to test the ability of AHFCAs to inhibit the binding of AvaL1 to DNA, reusing the SA-chip prepared in Chapter 4.2.1. AvaL1 was first premixed with AHFCA4; samples were diluted in SPR-SA Binding Buffer to final concentrations of 0.5 µM AvaL1 and 0.01 µM, 0.1 µM, 0.5 µM, 2.5 µM, 5M and 10 µM AHFCA4. The mixture was injected for 150 s at 30 µL min^{-1} to assay association, and dissociation of the complex was measured for 4 min. AHFCA4 inhibited the ability of AvaL1 to bind to all three immobilised AREs (Figure 6-3). The repeat assay (Figure 6-3 inset), showed a decrease in inhibition at low effector concentrations, but with similar inhibition at 10 µM AHFCA4, with all flow cells similarly affected. In the presence of AHFCA4 AvaL1 still exhibited multi-phasic binding in the sensorgrams, and almost all AvaL1 was inhibited from binding to the chip at higher AHFCA4 concentrations. Therefore, the non-1:1 Langmuir AvaL1 binding response is not as result of misfolded protein binding DNA (suggested as a possibility in Chapter 4.2.1), but because of the intrinsic mechanism of the protein interacting with DNA.

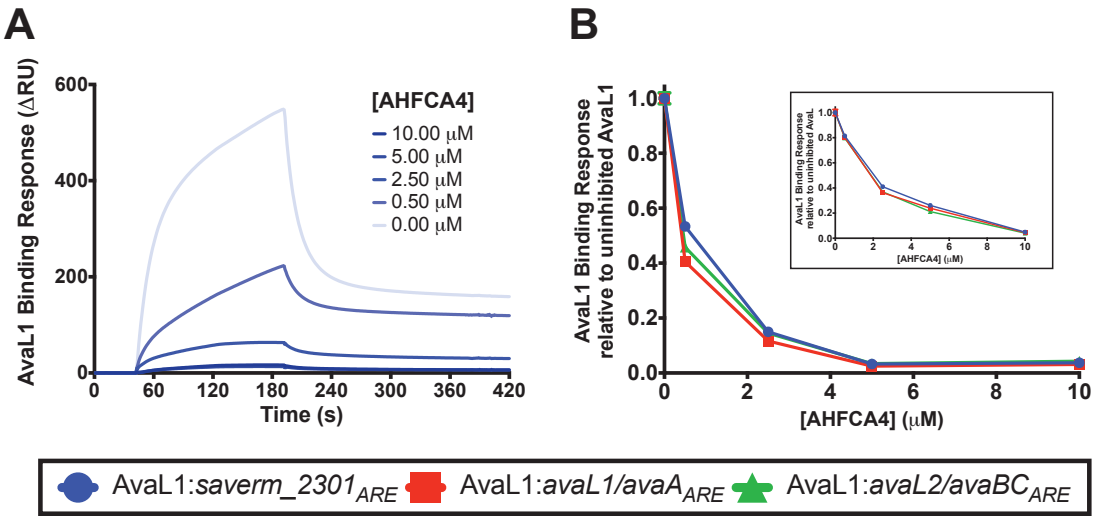


Figure 6-3 – Pilot Inhibition of AvaL1 binding to SA-immobilised *S. avermitilis* AREs in the presence of AHFCA4. 0.5 µM AvaL1 was incubated with 0.5 µM to 10 µM AHFCA4 prior to injection. Anomalous results for 0.01 µM and 0.1 µM AHFCA4 have been excluded from the graphs. (**A**) SPR reference-subtracted sensorgram for the binding of AvaL1 to *saverm_2301*$_{ARE}$. (**B**) shows the binding levels (*BL*; taken 4 s before dissociation phase) relative to uninhibited AvaL1 for all three immobilised DNA sequences. Inset graphs shows a repeat assay. Mean R_A and *BL* values are shown with 3 replicates for each concentration of AHFCA4, with SEM shown. The chip was reused from experiments described in Chapter 4.2.1.

IC_{50} curves were fitted to the data with the minimum and maximum response constrained to 0 and 1 relative AvaL1 binding respectively (not shown). Inhibition of AvaL1 binding to *saverm_2301*$_{ARE}$, *avaL1/avaA*$_{ARE}$ and *avaL2/avaBC*$_{ARE}$ by AHFCA4 had relative IC_{50} values of (0.50 ± 0.06) µM, (0.33 ± 0.02) µM, and (0.41 ± 0.03) µM respectively for the first pilot

experiment, and (1.68 ± 0.24) μM, (1.48 ± 0.21) μM and (1.44 ± 0.20) μM respectively for the second pilot experiment (which are consistent with later derived IC_{50} values). For both experiments the binding of AvaL1 to each ARE is homoscedastic within each group ($W(2, 12) = 0.02/< 0.01$, $p = 0.98/0.99$ for each experiment), and ANOVA indicates that the mean values are equal ($F(2, 12) < 0.01$, $p = 0.99$ for both experiments). For *in vivo* performed by Dr Styles AHFCA4 was ~3 times more effective at releasing MmfR bound to $mmfR/mmfLHP_{ARE}$ compared to MmfR bound to either $mmyB_{ARE}$ or $mmyR_{ARE}$, based on the resulting bioluminescence of the *lux* reporter system[149]. While there is some correlation between the K_D values of AvaL1 binding to each ARE (using data from Chapter 4.3.1) and the IC_{50} values of AHFCA4 ($r(1) = 0.93/0.52$ for each experiment), the statistical significance of this correlation is poor ($p = 0.24/0.65$ for each experiment). Therefore, there is no indicator otherwise that the ability of AHFCA4 (and presumably other AHFCA signalling molecules) is able to differentially prevent AvaL1 from binding to different AREs, and by assumptions of equilibrium release AvaL1 from different AREs. It is probable that promoter strength, as discussed in Chapter 4, influenced the amount of *in vivo* bioluminescence observed. Differences in the IC_{50} values between replicates most likely therefore result from disparities in sample dilutions, sample quality, degradation of the oligo fixed to the chip, *etc*.

6.2.2 Measuring relative inhibitory efficacy of a small library of effectors

Following determination of AHFCA activity in the pilot experiment, AvaL1 was premixed with the full set of effector molecules and injected over the SA-chip with $avaL1/avaA_{ARE}$ immobilised using 500 ng mL^{-1}, 50 ng mL^{-1} and 5 ng mL^{-1} oligonucleotide solutions, as described in Chapter 4.2.5. The experiment was repeated twice at different concentrations of effector to determine effector consistency across replicates, and the binding to all three flow cells where assessed (Figure 6-4). From visual inspection of the inhibitory activity of the effectors there are three distinct groups of activity; the first is comprised of the putative endogenous *S. avermitilis* effectors AHFCA(3, 5-7), which all exhibit strong inhibitory activity against AvaL1 DNA-binding. The second group consists of AHFCA(1, 4), which showed intermediate inhibitory activity against AvaL1, with AHFCA4 having the stronger activity of the two. The final group, comprised of SCB1, MmA and AHFCA2, all showed poor inhibitory activity against AvaL1, with AHFCA2 having the strongest effect and SCB1 the least.

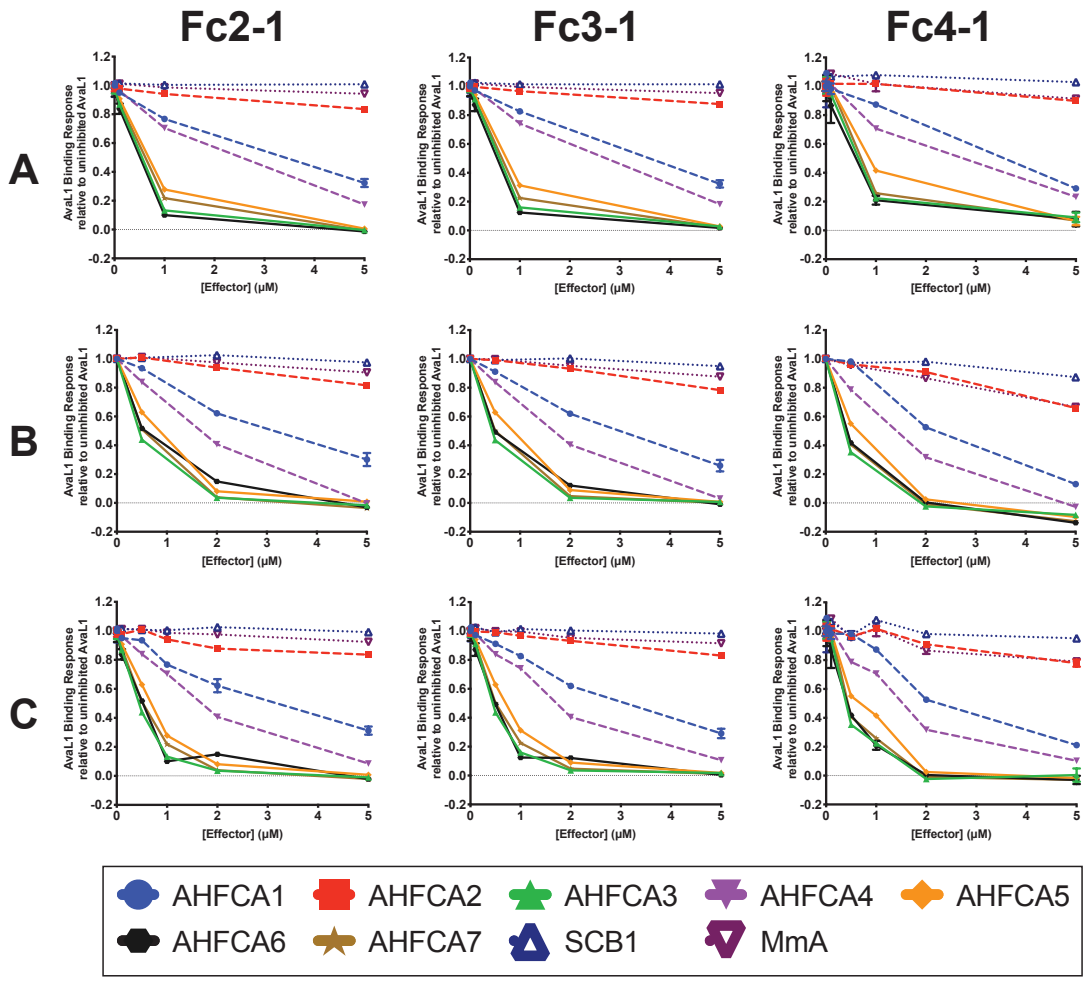


Figure 6-4 – Inhibition of AvaL1 binding to SA-immobilised *avaL1/avaA*$_{ARE}$ across three chip-surface densities by a small library of effectors. 0.5 μM AvaL1 was incubated with 0.01 μM to 5 μM of each effector prior to injection. Fc2 to 4 were immobilised with oligonucleotide solutions containing 500 ng mL^{-1}, 50 ng mL^{-1} and 5 ng mL^{-1} DNA respectively. Experimental protocols were the same as for the pilot inhibition experiment. (**A**) Relative AvaL1 *BL* values from first replicate set with effectors assayed at 0.01 μM, 0.1 μM, 1 μM and 5 μM. **B**) Relative AvaL1 *BL* values from the repeat assay set with effectors assayed at 0.5 μM, 2.5 μM and 5 μM. (**C**) Merged dataset of **A** and **B** for relative AvaL1. Putative *S. avermitilis* endogenous AHFCAs effectors are indicated with solid lines, exogenous AHFCAs with broken lines, and non-AHFCAs with dashed lines. Mean relative binding values are indicated with 3 replicates for each concentration, with SEM bars shown. The chip was reused from experiments described in Chapter 4.2.5.

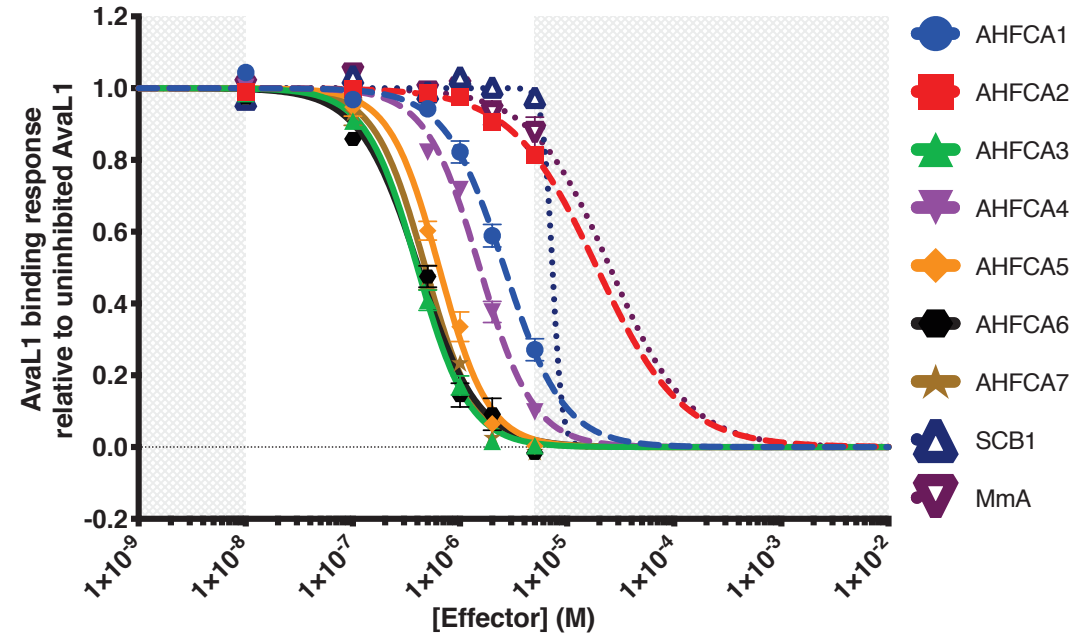
ANOVA of the mean relative AvaL1 binding response (for 3 replicates) for each flow cell indicates no significant difference between the three flow cells ($F(2, 186) = 0.01$, $p = 0.99$; with repeated experiments for each flow cell treated as a single group), and that the variance between cells was homoscedastic ($W(2, 186) = 0.24$, $p = 0.79$). Negligible differences between the AvaL1 binding responses on each flow cells (particularly where relative binding drops into negative values) could be account for by matrix effects, immobilised ligand density, ligand integrity, mass transfer effects, *etc*. As each flow cell corresponds to a different concentration of immobilised DNA, it can be concluded that the inhibitory effect (and presumed derepressive effect) of the effector is dependent only on the concentration of AvaL1 the effector binds, and not on the concentration of DNA.

6.2.3 Quantification of effector inhibitory activity

As there were no significant differences between the data from each flow cell, the datasets collected from Fc2-1, Fc3-1 and Fc4-1 were treated as replicates of one another and merged to produce a single dataset (FcMerged). Hill plots for the inhibitory activity for each effector (Figure 8-16, Appendix E) indicate cooperative inhibition of AvaL1 DNA-binding by all AHFCAs except for AHFCA2 (mean n_H value of 1.48 ± 0.45 excluding AHFCA2). The fits for SCB1 and MmA were not significant enough for accurate n_H determination ($p > 0.05$). A variable slope IC_{50} ($n_H \neq 1$) IC_{50} curve was fitted to the FcMerged data with an Ordinary Least-Squares regression, with the maximum and minimum relative AvaL1 response constrained to 1 and 0 respectively (Figure 6-5A). The derived IC_{50} values (Figure 6-5B) indicate that the three groups previously described can be grouped based on the order of magnitude of their determined IC_{50} values (Figure 6-5C), with values of 10^{-7} M for AHFCA(3, 5-7) (mean IC_{50} = 0.48 μM ± 0.06 μM), 10^{-6} M for AHFCA(1, 4) (mean IC_{50} = 2.06 μM ± 0.14 μM) and 10^{-5} M for AHFCA2 and MmA (mean IC_{50} = 22.08 μM ± 16.96 μM). For all AHFCA molecules the mean fitted n_H value was 1.63 ± 0.42, with AHFCA2 exhibiting the least positive cooperativity (n_H = 1.09 ± 0.12) and AHFCA5 the most (n_H = 1.88 ± 0.17), with a strong correlation between IC_{50} and n_H values ($r(5) = -0.91$, $p < 0.01$).

For SCB1 there was no correlation between ligand concentration and observed inhibitory activity ($r_s(61) = 0.01$, $p = 0.95$), and the IC_{50} curve fitting for SCB1 was poor ($R^2 = 0.16$). It can be concluded that the determined IC_{50} and n_H values for SCB1 are not accurate, and that SCB1 has no effect on the ability of AvaL1 to bind DNA; therefore, SCB1 can be considered a true negative control. By comparison, there was a significant correlation between ligand concentration and inhibitory activity for all other ligands ($p < 0.01$). Because the assayed range of relative *BL* values for AHFCA2 and MmA did not reach at least 50 % relative inhibition, the accuracy of the IC_{50} values for these ligands are debatable. Both AHFCA2 and MmA however have strong IC_{50} curve fits ($R^2 = 0.98$ and 0.86 respectively). Although there is lower correlation between effector concentration and AvaL1 activity for both AHFCA2 and MmA due to less inhibition being observed, the model accounts for much of the observed variance as indicated by the model R_s^2 values, so these IC_{50} values should be taken as strong predictions of inhibitory activity.

A



B

	Variable slope IC_{50} parameters			[Effector]-dependent inhibition		
Effector	**IC_{50} (μM)**	**n_H**	**R^2**	**r_s(61)**	**R_s^2**	**p**
AHFCA3	0.40 ± 0.02	1.80 ± 0.12	1	-0.97	0.94	< 0.01
AHFCA6	0.41 ± 0.03	1.56 ± 0.21	0.99	-0.98	0.96	< 0.01
AHFCA7	0.48 ± 0.03	1.83 ± 0.19	1	-0.95	0.91	< 0.01
AHFCA5	0.64 ± 0.04	1.88 ± 0.17	1	-0.97	0.95	< 0.01
AHFCA4	1.53 ± 0.08	1.73 ± 0.15	0.99	-0.94	0.88	< 0.01
AHFCA1	2.60 ± 0.12	1.57 ± 0.11	0.99	-0.95	0.90	< 0.01
AHFCA2	18.98 ± 3.38	1.09 ± 0.12	0.98	-0.75	0.57	< 0.01
MmA	25.18 ± 16.62	1.18 ± 0.42	0.86	-0.69	0.47	< 0.01
SCB1	7.30 ± 2159	9.66 ± 7559	0.16	0.01	0.00	0.95

C

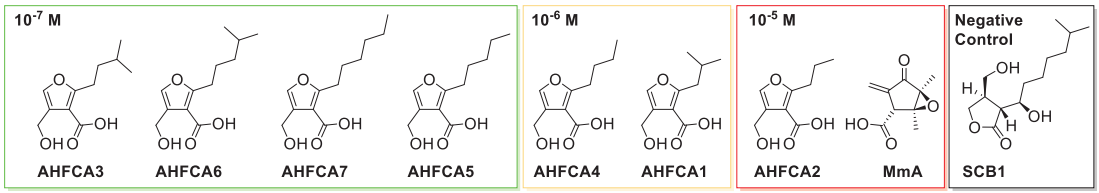


Figure 6-5 –IC_{50} analysis of inhibitory efficacy of small molecules in inhibiting the binding of AvaL1 to SA-immobilised *avaL1/avaA*$_{ARE}$. (**A**) Variable slope ($n_H \neq 1$) IC_{50} curves for the inhibition of AvaL1 binding to *avaL1/avaA*$_{ARE}$ using the data from the FcMerged dataset fitted with a constrained Ordinary Least-Squares regression. Fitting parameters were constrained to maximum relative response = 1 and minimum relative response = 0. The grey area of the graph indicates the projected regions. *S. avermitilis* endogenous AHFCAs are indicated with solid lines, exogenous AHFCAs with thick broken lines, and non-AHFCA ligands with smaller broken lines. Mean relative binding are shown for all three flow cells, with SEM shown. (**B**) Constrained relative IC_{50} and binding potential (*BP*) values for each effector as derived from **A**, colour-coded by IC_{50} value order of magnitude. *p* values are given for a Student's *t* test, with the null hypothesis that there is no correlation between the concentration of effector used and the relative amount of AvaL1 DNA-binding observed using replicate data sets. The nonparametric Spearman's correlation coefficient (r_s) was used to measure correlation as the relationship was non-linear. (**C**) Structures of assayed effectors grouped and ordered left to right by decreasing activity indicated by order of magnitude of the calculated IC_{50} value.

6.2.4 Effects of AHFCAs on DNA-binding by AvaL2

Dr Styles previously determined that MmyR does not transduce the signals of AHFCAs *in vivo* using a *lux* bioluminescent reporter system in *S. coelicolor* M145 (controlled through binding of MmfR/MmyR to AREs upstream of the synthetic *lux* operon)[149]. However, she also suggested that MmyR represses DNA in the presence of AHFCAs as evidenced by a reduction of luminescence in the presence of AHFCAs. Dr Sidda conversely noted a possible decrease in gaburedin production in an *S. venezuelae* Δ*gbnR* mutant at high induction concentrations of AHFCA[111], which may suggest that this change in repression levels is due to something other than the MmyR-type repressor. Dr Sidda attributed this decreased transcription of the *gbnABC* operon to SgnR silencing transcription as a result of overproduction, but the MmfR repressor is absent in Dr Styles' bioluminescence assays, so this change in activity may not be attributed to presence of the MmfR homologue.

To test whether an AHFCA signalling molecule could stimulate or inhibit AvaL2 binding to the *S. avermitilis* AREs bound by AvaL1, the initial SA Pilot assay chip immobilised with the three *S. avermitilis* azoxy compound gene cluster AREs as described in Chapter 4.2.1 was injected with 1 µM AvaL2 premixed with 0.01 µM to 20 µM AHFCA4 prior to injection. 1 µM AvaL1 (both isolated and AHFCA4 premixed) was also assayed as a reference for DNA-binding activity. AHFCA4 had no effect on the binding response of AvaL2 relative to non-induced AvaL2 references (Figure 6-6), which is reflected in a lack of significant correlation between AvaL2 *BL* values and AHFCA4 concentration ($r_s(4) = 0.23$, $p = 0.66$). AHFCA4 in the same experiments inhibited the binding of AvaL1 to all three of the immobilised AREs (not shown).

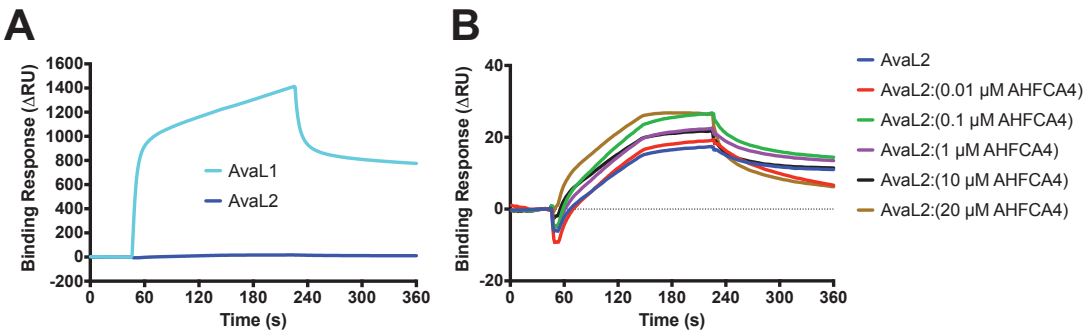


Figure 6-6 – SPR Sensorgrams showing lack of AHFCA4 activity against the DNA-binding activity of AvaL2. 0.01 µM to 20 µM AHFCA4 was premixed with 1 µM AvaL2 for 15 min and injected over an SA-chip coated with *saverm_2301*$_{ARE}$, *avaL1/avaA*$_{ARE}$ and *avaL2/avaBC*$_{ARE}$. (**A**) 1 µM Fc2-1 reference sensorgrams of AvaL1 and AvaL2 binding to *saverm_2301*$_{ARE}$. (**B**) Sensorgrams of mixed AvaL2:AHFCA4 solutions. Results were consistent across all three immobilised AREs. Mean R_A values are shown with 2 replicates for each concentration, with error too small to be displayed. For the example shown the Spearman's correlation between the concentration of AHFCA4 and the DNA-binding activity of AvaL2 was tested using a Student's *t* test ($r_s(4) = 0.23$, $p = 0.66$). The chip was reused from experiments described in Chapter 4.2.1.

6.2.5 Measuring AvaL1:ligand interactions using NTA-chip-based SPR

As the K_D value for an effector describes its binding affinity and IC_{50} describes its inhibitory activity, the two values for a single interaction may not always the same. For example, a competitive inhibitor that binds with a high affinity to a protein may not exhibit the same biological activity as the molecule with which it completes with. Secondly, the IC_{50} value is dependent on the concentration of available receptor/substrate, whereas K_D is not. Both terms are related by the Cheng-Prusoff equation: $K_I = \frac{IC_{50}}{1+([Effector]/K_D)}$, where K_I is the inhibition constant. If $IC_{50} \approx K_D$, then the affinity of binding should be proportional to the effectors ability to release/inhibit the repressors interaction with DNA.

To determine if this was the case for effectors binding to AvaL1, the direct binding of the effectors to AvaL1 was assayed with the His_6-tagged AvaL1 immobilised onto the chip as described in Chapter 2.4.12. After 200 nM of AvaL1 was injected over the chip surface, 2 μM solutions of AHFCA3 and SCB1 were injected, which both failed to produce a binding response. The recombinant AvaL1 dimer is 262-fold and 227-fold heavier than AHFCA3 and SCB1 respectively, therefore the effectors each produce a significantly smaller response compared to the binding observed for AvaL1 binding to the chip. Because the signal of AvaL1 binding to the chip was quite high (1000-1200 RU), the amount of effector injected was not enough to produce a strong enough signal. The solution for determining direct binding of effectors to AvaL1 may be to reduce the immobilisation level of AvaL1 on the chip surface.

The AvaL1 injected over the nickel-coated surface also dissociated from the chelated nickel because of unstable immobilisation to the chip surface (Figure 8-17, Appendix E). Either the His_6 tag was insufficiently strong enough to keep the protein immobilised on the surface, or the AvaL1 homodimer was disassembling at the chip surface. The former can be negated in future *via* incorporation of a longer polyhistidine tag (His_{10} or His_{12}) into the recombinant protein sequence. The latter issue could be tested for by introduction of a disulphide bridge between monomers, as is the case in the CprB repressor; however, this may have unintentional secondary effects on the function of the protein. Regardless, an unstable baseline meant the sensorgram was unable to be used to quantify kinetics, affinity or stoichiometry of binding.

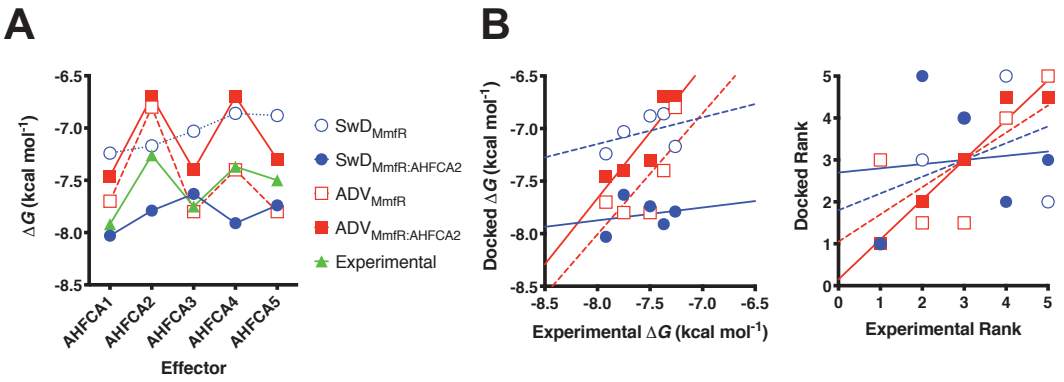
6.3 *In silico* modelling of MmfR:Effector interactions

6.3.1 Comparing docking methodologies for MmfR

Using AutoDock Vina (ADV) the endogenous *S. coelicolor* A3(2) AHFCA signalling molecules were docked to the crystal structures of both native MmfR and MmfR:AHFCA2 and correlated with the experimental data (Figure 6-7). Experimental ΔG at room temperature for each effector was determined using the equation $\Delta G = RT \ln K_D$, where K_D was assumed to be equal to the experimental IC_{50}, which was derived from the reciprocal of the binding potential (*BP*) values derived experimentally *in vivo* by Dr Styles[149] (and corroborated *in vitro* by Dr Zhou[148]). Previous docking of AHFCAs to the ligand-free structure of MmfR was performed by Dr Styles using SwissDock (SwD)[149]. For fair comparison between SwD and ADV, SwD was also used to dock AHFCA(1-5) to the MmfR:AHFCA2 structure. The derived order of binding for each method are as follows (with bold indicating a matched ranking to the experimental data):

$MmfR_{Experimental}$	**AHFCA1** > **AHFCA3** > **AHFCA5** > **AHFCA4** > **AHFCA2**
$MmfR_{SwD}$	**AHFCA1** > AHFCA2 > AHFCA3 > AHFCA5 > AHFCA4
$MmfR{:}AHFCA2_{SwD}$	**AHFCA1** > AHFCA4 > AHFCA2 > AHFCA5 > AHFCA3
$MmfR_{ADV}$	AHFCA3 = AHFCA5 > AHFCA1 > **AHFCA4** > **AHFCA2**
$MmfR{:}AHFCA2_{ADV}$	**AHFCA1** > **AHFCA3** > **AHFCA5** > **AHFCA4** = **AHFCA2**

ADV docking to the AHFCA2-bound MmfR structure produced the only ΔG values that ranked close to the observed *in vivo* order of affinity, with a significant correlation in experimental and simulated ΔG values by both parametric and nonparametric correlation ($r(3) = 0.90$, $p = 0.04$ and $r_s(3) = 0.98$, $p < 0.01$). ADV docking to the ligand-free MmfR structure however produced the closest range of mean docking scores to the mean experimental ΔG *values* (paired Student's $t(4) = 0.46$, $p = 0.67$). SwD docking to MmfR and MmfR:AHFCA2 overall by comparison had poor correlation with the *in vivo* data. ADV was the superior docking method and docking to the MmfR in the ligand bound conformation produces stronger predictions of affinity. Assuming the ADV docking is a good representation of the binding affinity, then it can be inferred for this group of ligands that the IC_{50} value has a high correlation of activity with the binding K_D value.



C

Effector	SwD_{MmfR}	$SwD_{MmfR:AHFCA2}$	ADV_{MmfR}	$ADV_{MmfR:AHFCA2}$	*in vivo*
	ΔG (kcal mol^{-1})				
AHFCA1	-7.24	-8.03	-7.70	-7.46	-7.92
AHFCA2	-7.17	-7.79	-6.80	-6.70	-7.26
AHFCA3	-7.03	-7.63	-7.80	-7.40	-7.75
AHFCA4	-6.86	-7.91	-7.40	-6.70	-7.37
AHFCA5	-6.88	-7.74	-7.80	-7.30	-7.50
Mean	-7.04 ± 0.18	-7.82 ± 0.16	-7.50 ± 0.42	-7.11 ± 0.38	-7.56 ± 0.27
	Pearson's correlation (parametric)				
***r*(3)**	0.40	0.21	0.74	**0.90**	-
p	0.50	0.74	0.15	**0.04**	-
	Spearman's correlation (nonparametric)				
r_s(3)	0.40	0.10	0.67	**0.98**	-
p	0.51	0.87	0.22	**< 0.01**	-

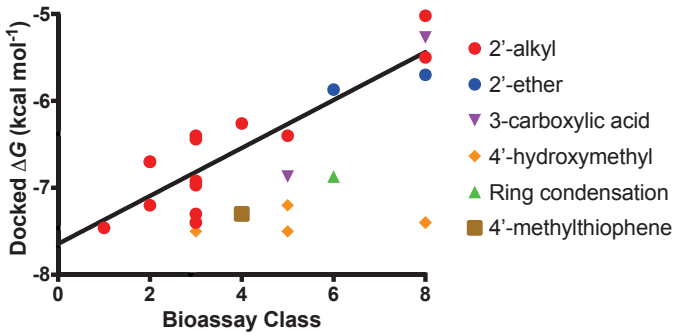
Figure 6-7 – Correlation of estimated *in vivo* Δ*G* values for MmfR activity inhibition by AHFCA(1-5) with docking simulations performed using SwissDock (SwD) and AutoDock Vina (ADV). (**A**) Plot of experimental and simulated Δ*G* values. (**B**) Plots of (left) docked Δ*G* values against experimental Δ*G* values and (right) docked rank against experimental rank, both fitted with a linear regression line. (**C**) Table of Δ*G* values with Pearson's and Spearman's correlation statistics. The mean for each set of Δ*G* values are shown with standard deviation. Correlation statistics for each docking dataset are indicated between that docking dataset and the *in vivo* dataset. Δ*G* values for the experimental data collected by Dr Styles[149] are estimated by equivalence of ligand affinity (K_D) with ligand inhibitory activity (IC_{50}), with IC_{50} calculated from the reciprocal of the binding potential (*BP*) for each ligand.

6.3.2 Docking synthetic AHFCA analogues to MmfR

Docking was also performed for the synthetic AHFCAs (based on the core structure of AHFCA1) screened for MmA bioassay activity by Dr Malet[139]. The *in vivo* data showed that none of the synthetic analogues improved activity over AHFCA1. AHFCA molecules with very small and very large alkyl chains produced a weaker response. The removal, substitution or condensation of the carboxylic acid moiety greatly reduced activity, while removal or substitution of the hydroxymethyl moiety also showed reduced activity to a lesser degree. Addition of an ether group between the terminal carbon and the adjacent carbon into the alkyl chain also reduced activity, indicating hydrophobicity of the alkyl moiety is also important factor. The substitution of the 4-hydroxymethylfuran core with a 4-methylthiophene core had slightly lower affinity than the 4-methylfuran analogue, indicating that electronegativity of oxygen drives better binding compared with the less electronegative sulfur.

Each effector from these experiments were grouped into bioassay classes based on the minimum amount required to induce a response; from AHFCA1 at rank 1 requiring only 0.05 µg to induce activity, down to rank 8 for compounds which did not induce a response. Each molecule was also grouped based on the modifications made to the core AHFCA structure. There is weak positive correlation between the docked scores and the experimental bioassay classes for all 25 compounds had an ($r(23) = 0.52$, $p < 0.01$) (Figure 6-8), with the docking of 2'-alkyl + 2'-ether modified compounds having the strongest correlation ($r(14) = 0.89$, $p < 0.01$), and the docking of 2'-alkyl + 2'-ether + 3'-carboxylic acid modified compounds being the most significant ($r(16) = 0.89$, $p < 0.01$). Using the most significant correlation as a reference line, the 4'-methylthiophene and condensed ring structures docked slightly better than the experimental values, while the 4'-hydroxylmethyl modified structures docked better by a significant margin. Of the two structures that scored better that AHFCA1, both belonged to 4'-hydroxylmethyl modified structures (compounds 131 and 134), and except for compound 132 all MeOH substitutions scored better than AHFCA3. The docking scores of AHFCA molecules with 4'-hydroxylmethyl substitutions did not correlate well with the experimental results ($r(3) = 0.19$, $p = 0.76$).

A



B

Type (Based on modification)	*n*	*r*(*n*-2)	R^2	*p*
All	25	0.53	0.28	< 0.01
2'-alkyl	14	0.88	0.77	< 0.01
2'-alkyl + 2'-ether	16	0.89	0.79	< 0.01
2'-alkyl + 2'-ether + 3'-carboxylic acid	18	0.89	0.78	< 0.01
2'-alkyl + 2'-ether + ring condensation + 4'-methylthiophene	18	0.81	0.66	< 0.01
2'-alkyl + 2'-ether + 4'-hydroxymethyl	21	0.47	0.22	0.03
4'-hydroxymethyl	5	0.19	0.04	0.76
All except for 4'-hydroxymethyl	20	0.82	0.67	< 0.01

Figure 6-8 – Correlation of *in vivo* methylenomycin bioassay results using AHFCA synthetic analogues with ADV docking scores. (**A**) Plot of docking scores against each compounds bioassay class, fitted with a linear regression through data of the 2'-alkyl, 2'-ether and 3'-carboxylic acid groups. Docking values for synthetic AHFCAs can be found in the appendices. (**B**) Table of correlation statistics of docking data with different groups of AHFCA homologues. Docking scores for each compound are shown in Figure 8-20, Appendix E. *n* denotes the number of compounds within the group.

AHFCA 4'-hydroxymethyl modified structures substituted with heteroatom containing groups (compounds 133, 134 and 152) all formed interactions with H84 in the docking simulations. However, compounds 131 and 132 both lacked heteroatom moieties to form this interaction with H84, so hydrogen bonding alone cannot account for all the docking values of the 4'-hydroxymethyl substituted structures. Although the hydroxymethyl moiety of the AHFCA2 structure does not participate in electrostatic interactions with H84 in the MmfR:AHFCA2 crystal structure, its presence still appears to improve AHFCA recognition. This moiety may infer additional polarity, and therefore possibly improve the solubility for molecules. Using MarvinSketch, the intrinsic solubility (*logS*) of AHFCA1 and 4'-hydroxymethyl modified analogues were predicted and plotted with their activity (Figure 8-19, Appendix E), but no significant correlation was observed between the two variables ($r(4) = -0.10$, $p = 0.86$). Other limiting factors of *in vivo* assays such as cell membrane permeability or media diffusion may account for a decrease in observed activity. To conclude, estimating affinity of AHFCA analogues *in silico* may have some value, based on strong correlation of the available *in vivo* and *in silico* data for some AHFCA analogues. However, docking for particular structures, such as 4'-hydroxymethyl substituted analogues, has limited predictive power.

6.3.3 Modelling affinity of AHFCAs for other MmfR orthologues

To explore the tolerance of different length alkyl moieties in AHFCA compounds, a library of virtual compounds (PDB structures available in the Electronic Supplementary Materials) was generated and docked with ADV to the structure of MmfR:AHFCA2. AHFCA structures containing *n*-alkyl, isoalkyl and *tert*-alkyl moieties were generated to investigate the effects of methyl branching at the terminus of the alkyl chain for each value of *x*, where *x* is defined as the alkyl length in the longest chain of unbranched carbon atoms (AHFCA alkyl moieties listed in Table 8-10, Appendix E). If $x = 2$, this is an ethyl group for an *n*-alkyl linear chain, isopropyl (1-methylethyl) for a methyl-branched chain, and a *tert*-butyl (1,1-dimethylethyl) for a dimethyl-branched chain, *etc*. Because of the number of possible structural isomers, structures containing branched methyl groups connected at non-preterminal carbons were not generated. In addition to docking to the MmfR:AHFCA2 crystal structure, homology models of AvaL1, SgnR and SclM1 were generated with Phyre[2] using local alignment of the MmfR:AHFCA2 structure as a template.

A distinctive "troughing" of docking scores is seen for a linear alkyl chain, where the ΔG values decreases from *x*0 (compound 70; HFCA) through *x*5 (AHFCA5) and then increases up to *x*12 (Figure 6-9). Single methyl branching improved docking scores only for short alkyl chains of *x*2-4, with the lowest ΔG value at *x*3 (AHFCA1) and the isoalkyl branched structure

*x*5 having a similar Δ*G* value to the n-alkyl structure (branched AHFCA6 and linear AHFCA5). Dimethyl branching generated universally poor docking scores compared to *n*-alkyl and isoalkyl AHFCAs of the same value of *x*. AHFCA structures with *x*1-9 for *n*-alkyl chains, *x*2-7 for isoalkyl chains and *x*2-4 for *tert*-alkyl chains all docked better than HFCA, with *x*6 *n*-alkyl (AHFCA7), *x*2 isoalkyl (compound 124) and *x*5 isoalkyl (AHFCA6) being the only nonendogenous AHFCA structures which scored better than AHFCA2, with order of docking scores indicated as thus:

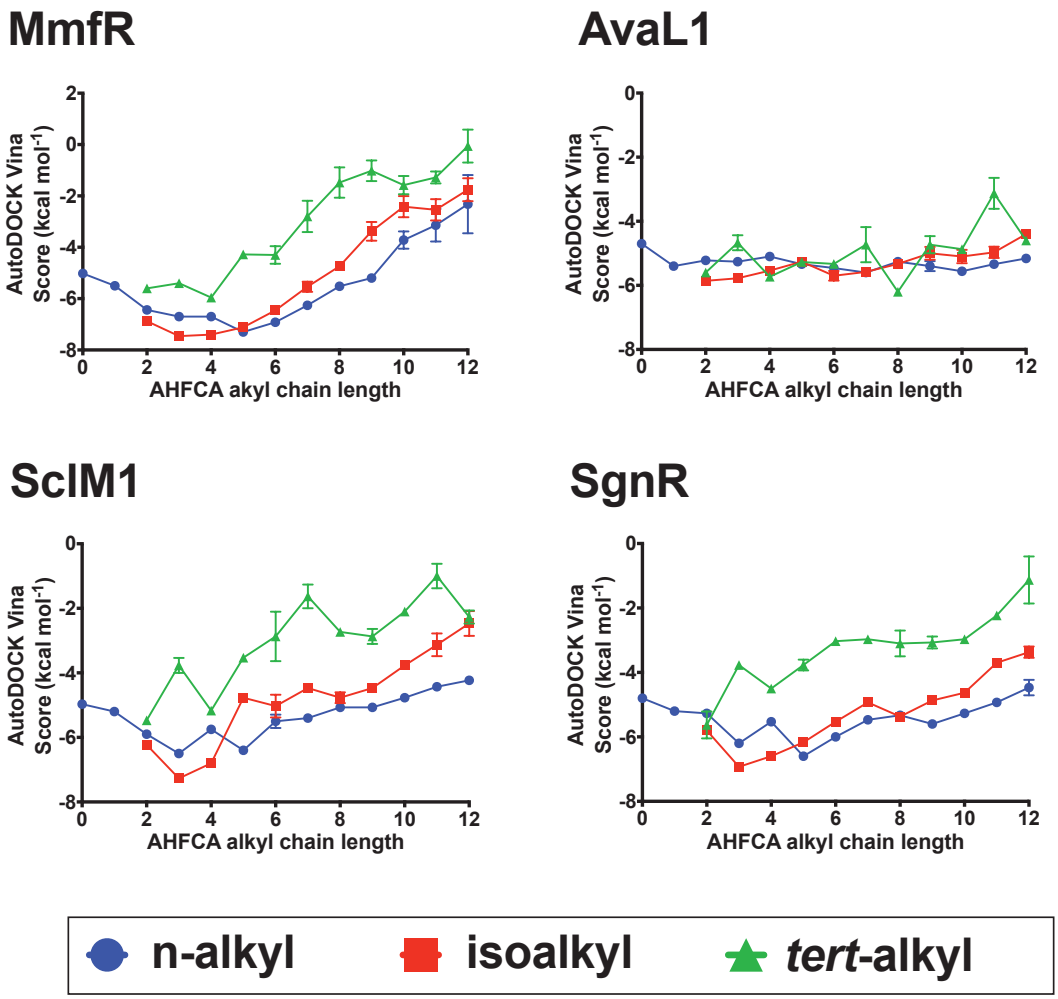
AHFCA1 > AHFCA3 > AHFCA5 > AHFCA6 > AHFCA7 > 124 > AHFCA4 = AHFCA2

This indicates that *S. coelicolor* A3(2) produces many of the most optimal length/branched AHFCA effectors for activity against MmfR, and that MmfR has a preference for AHFCAs utilising precursors produced in the *S. coelicolor* A3(2) fatty acid metabolism pathway. MmfR may also have greater tolerance for some of the larger AHFCAs than previously anticipated. The highly conserved LBPs in SclM1 and SgnR lead to a similar consensus of docking patterns for SclM1 ($r(103) = 0.87$, $p < 0.01$) and SgnR ($r(103) = 0.87$, $p < 0.01$), but with a decrease in docking scores. The LBP of the SgnR Phyre^2 model has a solvent-accessible area of 371 Å^2, and the SclM1 Phyre^2 model 402 Å^2; a respective 1.55-fold and 1.68-fold increase over the size of the MmfR:AHFCA2 LBP. This increase is not consistent with the expected surface area based on direct substitutions of residues in PyMOL (245 Å^2 and 254 Å^2 for SgnR and SclM1 respectively), compared with 239 Å^2 for the MmfR:AHFCA2 LBP. This is possibly due to minimisation of the structures during the model generation process. The number of structures that scored better than HFCA was 15 in SclM1 and 20 in SgnR, compared with 18 in MmfR. The data trends suggest that AHFCA2 performed better than expected in both SclM1 and SgnR models, so instead AHFCA4 was used as a baseline (assuming equivalence to AHFCA2, to which it scored equally in MmfR docking). The order of docking scores are as follows:

SclM1 AHFCA1 > AHFCA3 > 124 > AHFCA5 > 120 > AHFCA4 ≈ AHFCA2
SgnR AHFCA1 > AHFCA3 = AHFCA5 > AHFCA6 > AHFCA7 > 124 > AHFCA8 = AHFCA4 ≈ AHFCA2

This may suggest that both SclM1 and SgnR have a similar order of affinity to MmfR, with SgnR being more tolerant to larger AHFCAs, such as those produced by *S. venezuelae*. It is not certain how well this docking data correlates with the ligand specificity of SgnR for AHFCAs produced in *S. venezuelae* and SclM1 for AHFCAs produced *S. sclerotialus*. The data could be rationalised with corroborating *in vitro* data based on the protocols outlined in this chapter.

A



B

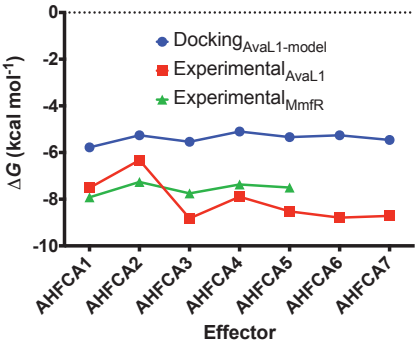
	AvaL1	SclM1	SgnR
All AHFCA molecules			
r(103)	0.53	0.87	0.87
p	< 0.01	< 0.01	< 0.01
n-alkyl AHFCA molecules			
r(37)	0.07	0.90	0.80
p	0.67	< 0.01	< 0.01
Isoalkyl AHFCA molecules			
r(31)	0.86	0.88	0.93
p	< 0.01	< 0.01	< 0.01
***tert*-alkyl AHFCA molecules**			
r(31)	0.42	0.79	0.84
p	0.02	< 0.01	< 0.01

Figure 6-9 – ADV scores for the *in silico* docking of an AHFCA virtual analogue library generated in MarvinSketch to MmfR:AHFCA2 and MmfR:AHFCA2 Phyre² homology models of AvaL1, SclM1 and SgnR. (**A**) Plotted docking scores for each AHFCA receptor, show mean ΔG values with SEM shown for 3 replicates based on selection of AHFCA conformations forming interactions via their carboxylic acid moieties to Y85 and Y144. Table of structures can be found in Table 8-10 and docking results in Table 8-11 through Table 8-14, Appendix E. (B) Pearson's correlation between the mean ligand docking scores for MmfR and AvaL1/SclM1/SgnR.

The Phyre² modelled AvaL1 structure has a solvent-accessible area of 462 Å², with two possible entrances to the LBP. Compared with the docking scores of the other MmfR orthologues, ADV docking of AHFCAs to AvaL1 produced a set docking scores which, while having a significant overall (but poorer) correlation ($r(103) = 0.53$, $p < 0.01$), also had poor correlation for n-alkyl ($r(37) = 0.07$, $p = 0.67$)-chained AHFCAs. All but four of the structures scored better than HFCA, while 21 scored better than AHFCA2. Single methyl branching

improved docking scores for *x*2-4 and *x*6 while *tert*-alkyl scores compared with *n*-alkyl and isoalkyl structures are erratic to say the least. These docking scores were correlated with the *in vitro* SPR data by estimating the Δ*G* for binding to AvaL1 as done for the *in vivo* activity against MmfR (Figure 6-10); while docking scores for AvaL1 correlated with MmfR docking scores, there doesn't appear to be any predictive power of the AvaL1 docking scores when compared with the experimental Δ*G* values by either mean Δ*G* values ($r(5) = 0.08$, $p = 0.86$) or by ranking ($r_s(5) = 0.22$, $p = 0.64$). It is clear that the positions of amino acid residues in the LBP of AvaL1 cannot be modelled one-to-one with the crystal structure of MmfR. Without a crystal structure of AvaL1, it is impossible to derive any meaningful *in silico* data about this repressor.

A



B

	ΔG (kcal mol^{-1})	
Effector	**ADV$_{AvaL1\text{-}model}$**	***In vitro***
AHFCA1	-5.78	-7.61
AHFCA2	-5.26	-6.44
AHFCA3	-5.54	-8.72
AHFCA4	-5.10	-7.93
AHFCA5	-5.34	-8.44
AHFCA6	-5.26	-8.71
AHFCA7	-5.46	-8.62
	Pearson's	**Spearman's**
***r*(5)**	0.08	0.22
p	0.83	0.64

Figure 6-10 – Correlation between ADV scores for AHFCA effectors docking to the MmfR:AHFCA2 Phyre^2 AvaL1 homology model and the SPR-derived activity against AvaL1. (**A**) Plot of docking scores for each ligand, with experimental data for MmfR determined as for Figure 6-7. Shown are SEM for 3 docking replicates selected for optimum orientation in the modelled LBP. *p* values are given for both parametric (Pearson's) and nonparametric (Spearman's) correlations between docked and experimental docking data.

6.3.4 Designing AHFCA analogues *in silico*

To explore the concept of structure-led design for a novel effector of MmfR, an iterative structure-by-structure method to improve upon the docking score of AHFCA1 was employed. Each compound, as before, was docked against the MmfR:AHFCA2 using ADV (Figure 6-11). As the activity of variant structures has already been explored *in vivo* by Dr Malet[139], modifications which had not been previously characterised were investigated. Firstly, the furan ring itself was manipulated, which has only been altered in the 4'-methylthiophene compound 158. Decreasing the number of carbon atoms to three (forming a 2H-oxete ring with two enantiomers; cAJF1s and cAJF1r) and increasing the number of carbon atoms to five (2H-pyran; cAJF2) both resulted in a poorer docking score. Substitution of the furan ring with (1H-pyrrol-1-yl)methanol (cAJF3) and 1H-pyrrole-1-carbaldehyde (cAJF4) also resulted in poorer docking. Substitution of the furan ring with a cyclopenta-2,4-dien-1-one ring (cAJF5) however improved the docking score by 0.4 kcal mol^{-1}.

Another component considered was the rearrangement of moieties around the furan ring. Based on AHFCA1, the 2-alkyl, 4-hydroxymethyl and 3-carboxylic acid moieties were shunted one carbon around the ring (cAJF6; 3-isobutyl-5-hydroxylmethylfuran-4-carboxylic acid) which improved the docking score by 0.2 kcal mol^{-1}. cAJF6 had an almost identical conformation as AHFCA1 in the binding pocket, the difference being the formation of a new interaction between the oxygen in the furan ring and side-chain of Q183. Substituting the ring of this shunted structure with a cyclopenta-2,4-dien-1-one ring (replacing furan ether with ketone; cAJF7) scored worse than cAJF6 by 1.3 kcal mol^{-1}.

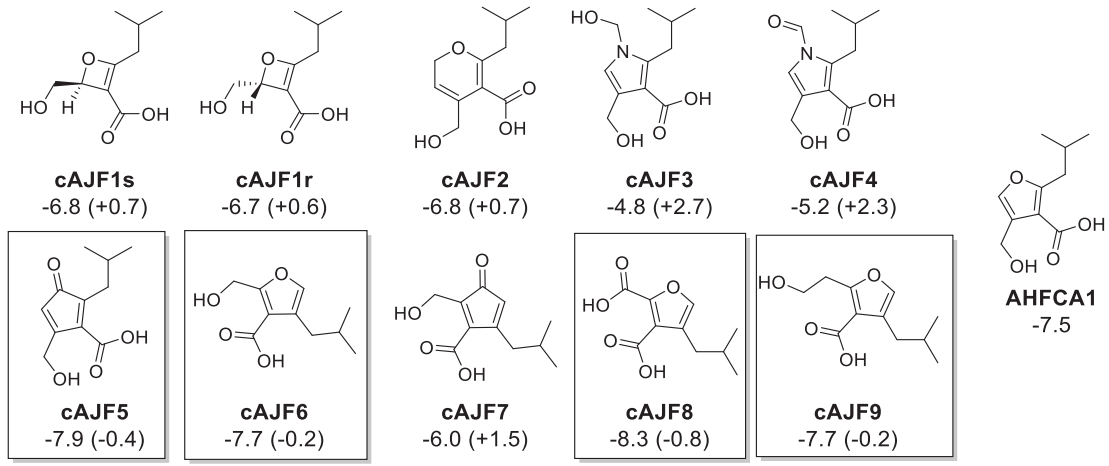


Figure 6-11 – ADV Docking of novel molecules generated iteratively based on the core structure of AHFCA1 to the MmfR:AHFCA2 crystal structure. Structures which scored better than AHFCA1 are enclosed in a box. Docking score are indicated in kcal mol^{-1}, with values in parenthesis show differences in docking score from AHFCA1. Conformations of each structure in the MmfR LBP are shown in Figure 8-20, appendix E.

Finally, the hydroxymethyl moiety was manipulated. Substitution with a hydroxyethyl moiety (cAJF9) resulted in no improvement over cAJF6, but substitution with a carboxylic acid led to the final docked compound, 4-isobutylfuran-2,3-dicarboxylic acid (cAJF8), which improved the score over cAJF6 by 0.6 kcal mol^{-1}, for a final docking score of -8.3 kcal mol^{-1}. This cAJF8 docking score is equal to a theoretical K_D value of 0.78 μM at 25 °C compared with the docked K_D value of 3.00 μM and a putative experimental K_D value (based on *BP* value) of 1.49 μM for AHFCA1. Projecting the difference in the experimental and docked K_D values for AHFCA1 onto the cAJF8 structure, cAJF8 had a predicted IC_{50} value of 0.38 μM. cAJF8 formed four electrostatic interactions in the LBP; the key interactions with Y85 and Y144 were formed with the 3-carboxylic acid, the 2-carboxylic acid interacted with H84 and the oxygen of the furan ring interacted with the amide group of the Q189 sidechain. However, because of this interaction with H84, it is possible, like observed for docking of Dr Malet's AHFCA analogues, that the structure has been over-docked to the binding pocket of LBP. In which case, cAJF6 might present the best reasonable alternative.

The literature suggests that furan-2,3-dicarboxylic acid has similar properties to AHFCA. Both present similar health and safety considerations including skin irritation, serious eye irritation and respiratory irritation. Both molecules are also hydrophobic but both furan-2,3-dicarboxylic acid and its structural homologue furan-2,5-dicarboxylic acid are soluble in DMSO. A possible synthetic route to cAJF8 has already been published[239, 240]. It is highly possible though that the optimum effector has already been described, although mutagenesis of the LBP may also lead to future discovery of further novel effectors.

6.4 Characterising the AHFCA structure-activity relationship

6.4.1 Comparing AHFCA activity between MmfR and AvaL1

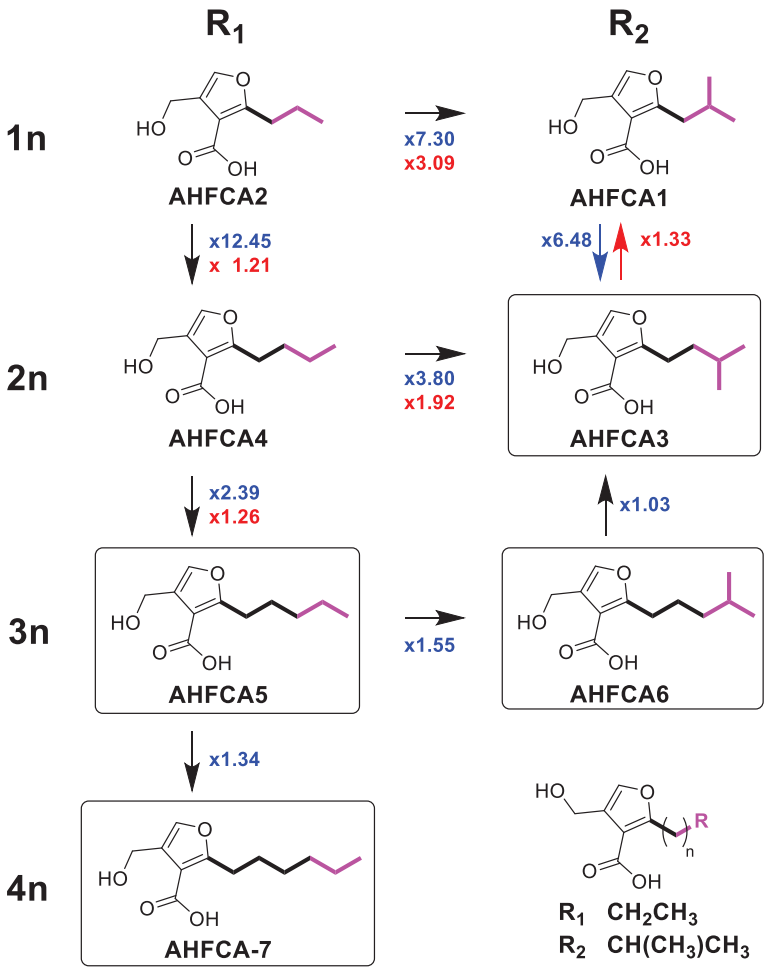
The order of AHFCA activity against both MmfR and AvaL1 are as follows:

MmfR AHFCA1 > AHFCA3 > AHFCA5 > AHFCA4 > AHFCA2
AvaL1 AHFCA3 > AHFCA6 > AHFCA7 > AHFCA5 > AHFCA4 > AHFCA1 > AHFCA2

To compare their activity more directly, *BP* values for each AHFCA molecule were compared, and these values correlated with the structure of each AHFCA molecule (Figure 6-12). As IC_{50} values were determined using the *BP* values for MmfR, the *BP* values for AvaL1 were determined from the reciprocal of the IC_{50} values. For AHFCA(1-5) there was no significant difference in mean *BP* values between MmfR and AvaL1 (paired Student's $t(4) = 1.45$, $p = 0.22$); although the variance in *BP* values is greater for AvaL1 (standard deviations of 0.19 μM^{-1} and 0.99 μM^{-1} for MmfR and AvaL1 respectively; this can be visualised in the ΔG values shown in Figure 6-10). One possible cause of the difference in variance may arise from the two different methodologies used to assay AHFCA activity (*in vivo* and *in vitro*). Because of the many factors that have to be accounted for in interpreting the *in vivo* data, such as rate-limited diffusion through culture media, differences in cell membrane permeability, catabolism, *etc*, *in vitro BP* values may different than those observed *in vivo*. Replicating MmfR AHFCA interactions using SPR would produce the most consistent comparison.

Another explanation may be a difference in the recognition of AHFCA signalling molecules by these repressors. In the case of all AHFCAs binding to MmfR and AvaL1 the specific inhibitory activity of each effector is attributed to the hydrophobic alkyl moiety. AHFCA2, the smallest *Streptomyces* spp. produced AHFCA molecule, has the shortest alkyl chain and demonstrates the weakest activity against both AvaL1 and MmfR. Work by Dr Styles has shown that AHFCA lacking the alkyl chain (HFCA) has an extremely poor ability to release MmfR from DNA *in vivo*[149]. Similarly, Dr Malet has also demonstrated *in vivo* that both

HFCA and compound 80 (2-methyl) do not induce MmA production in *S. coelicolor* A3(2), while compound 120 (2-ethyl) and the same-length branched compound 124 (2-isopropyl) were increasingly able to induce MmA production, but less so than the endogenous *S. coelicolor* A3(2) AHFCA molecules[139].



BP (μM^{-1})	AHFCA1	AHFCA2	AHFCA3	AHFCA4	AHFCA5	AHFCA6	AHFCA7
MmfR	0.67	0.22	0.51	0.26	0.33	ND	ND
AvaL1	0.38 ± 0.02	0.05 ± 0.01	2.49 ± 0.11	0.66 ± 0.03	1.57 ± 0.07	2.43 ± 0.17	2.10 ± 0.12

Figure 6-12 – Structures of effectors annotated with the observed difference in their efficacy against MmfR and AvaL1 DNA-binding activity. Fold-increases in activity are in **red** for the *in vivo* deregulation of MmfR (as determined by Dr Styles[149]) and **blue** for the *in vitro* inhibition of AvaL1, with arrows in the direction of increasing efficacy. AHFCAs putatively produced by *S. avermitilis* are enclosed in rectangular boxes. ND = not determined. AvaL1 *BP* values are derived from reciprocal IC_{50} values determined in Figure 6-5.

The experimental data indicates that for MmfR the longer the AHFCA alkyl chain becomes the stronger the AHFCAs inhibitory activity, with the *in silico* docking data indicating binding improves up to a given length, with diminishing returns in the fold-increase in activity as alkyl chain length increases. Methyl branching of the alkyl chain for a given length of the alkyl chain also increases the activity of each AHFCA over its same length linear counterpart (AHFCA1 > AHFCA2, AHFCA3 > AHFCA4, *etc*); again, with diminishing returns. The *in*

silico data also indicate that as the length of the alkyl chain continues to increase the affinity of the AHFCA molecule for the target receptor will eventually start to decrease. This is also corroborated in the *in vivo* work performed by Dr Malet, which indicates that AHFCA8E (2-heptyl) has a weaker ability to induce MmA production than AHFCA7 (2-hexyl) and the five endogenous *S. coelicolor* A3(2) AHFCAs.

The same trends are reflected in the experimental data for AvaL1. These trends indicate that there is an optimum configuration of alkyl chain length and branching for each AHFCA receptor; these are AHFCA1 for MmfR and AHFCA3 for AvaL1, both of which are produced by their endogenous AHFCA biosynthetic systems. AHFCA1 has an alkyl chain that is one carbon shorter than AHFCA3, and there is a much greater disparity between the larger AHFCA molecules, which are more effective against AvaL1 than MmfR, and the smaller molecules, which are less effective against AvaL1 than MmfR. It can be inferred that AvaL1 is more receptive to larger AHFCAs compared with MmfR, correlating with the observed AHFCA molecules produced by *S. avermitilis*, which have generally larger alkyl chains than those produced by *S. coelicolor* A3(2).

6.4.2 Rationalising differences in AHFCA response *in silico*

The residues contributing to the LBP of MmfR are H84, **Y85**, L110, A113, F117, M123, A127, **Q130**, P143, **Y144**, W147, V178, F181, F182, and Q185 (key interaction residues in bold); for simplicity these residues can be written as H**Y**LAFMA**Q**P**Y**WVFFQ. AvaL1 has the least conserved LBP of the known AHFCA receptors (F**Y**MLFIA**Q**P**Y**TGVFQ). The key residues for AHFCA binding, Y85 and Y144, as well as Q130 are conserved across all confirmed AHFCAs receptors. H84 is substituted with phenylalanine in AvaL1 (F89); while H84 of MmfR is in proximity to the AHFCA hydroxymethyl moiety, the absence of interactions formed in the MmfR:AHFCA2 crystal structure and its absence in AvaL1 suggest that this residue does not play a significant role in forming electrostatic interactions with AHFCA molecules. H84 contributes towards the pockets geometry but may also function in some MmfR orthologues to provide electrostatic orientation for AHFCAs.

In AvaL1 most residues with hydrophobic side-chains are substituted with similarly hydrophobic residues. Two of these substitutions; W147 with threonine (T153 in AvaL1) and F181 with valine (V187 in AvaL1) results in an expansion the solvent-accessible hydrophobic region in proximity of the AHFCA hydrophobic alkyl chain moiety in a residue-mutated PyMOL *in silico* model of AvaL1 based on the structure of MmfR (Figure 6-13). The LBP of MmfR with the residues directly mutated has a larger solvent-accessible area of 450 Å^2

(dependant on selected amino acid rotamers), compared with 263 Å^2, and 239 Å^2 for ligand-free and AHFCA2-bound MmfR respectively. However, it is probable that the geometry of the LBP is not as shown in this model (which was used because of poor docking with the Phyre^2 homology model).

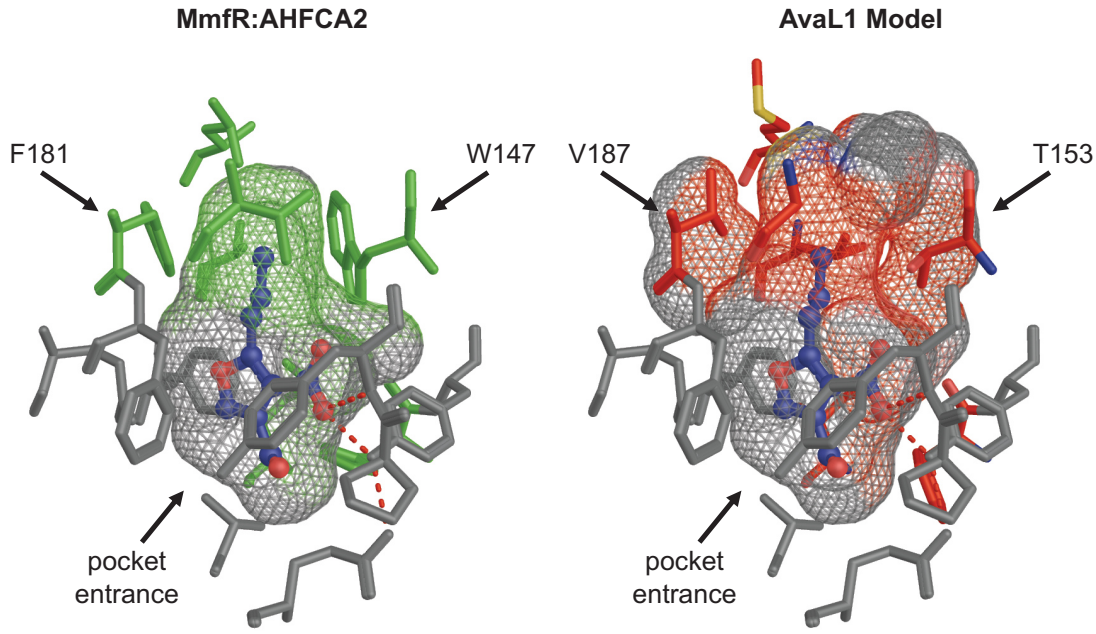


Figure 6-13 – Comparison of solvent accessible cavities of the LBP of the MmfR:AHFCA2 crystal structure and a PyMOL homology model of AvaL1 based on the MmfR:AHFCA2 crystal structure. Grey residues are conserved in both repressors, while green residues are those found only in MmfR and red those found only in AvaL1. LigPlot of the MmfR structure can be found in Figure 1-20, Chapter 1.4.4.

The substitution of the tryptophan residue is itself atypical, as this residue is well conserved in many other ArpA-subfamily repressors including CprA, CprB, BarA, FarA, ScbR, ScbR2, AvaR1, AvaR3, JadR3, BulR1, BulR2 and TylP, as well as MmfR, SgnR and SclM1. A W119A mutation in ArpA has previously been shown to abolish A-factor response[(235)], possibly implying this residue is important in GBL binding across the ArpA-subfamily. However, it is not found in either of the pseudo-GBL receptors AvaR2 or JadR2, nor in any of the pseudo-AHFCA receptors. The absence of this tryptophan in AvaL1 suggests it is not necessary for AHFCA binding in AvaL1, but whether this tryptophan is necessary for AHFCA binding in other AHFCA receptors is not known.

The docking data indicates that because of the small size of the LBP for the AHFCA receptors characterised, increases in the length of the AHFCA alkyl chain will only increase the ligand-binding affinity up to a certain length; above which the affinity will decrease because of the less energetically favourable conformations that need to be assumed by the AHFCA molecule to form the key interactions *via* Y85 and Y144. The trends also indicate that for AHFCAs with

longer alkyl chains the addition of methyl branches will decrease the affinity of the molecule rather than increase it, for the same reasons. The effect of these two substitutions in AvaL1 on effector binding is two-fold; first, as the hydrophobic region in the LBP is expanded, longer chain AHFCA molecules can more rigidly fit into the pocket without exceeding the maximum capacity of the pocket. Secondly, as smaller alkyl chain AHFCAs no longer fully occupy the available space, they experience greater 'wobbling' of the alkyl chain. This model corroborates the experimental data, and accounts for the larger variance in AHFCA activity.

Both Dr Styles and Dr Zhou have previously shown that SCB1 failed to release the MmfR repressor from DNA both *in vivo* and *in vitro*[148, 149], and here it is reported that SCB1 also has no activity against AvaL1. Although SCB1 has the same alkyl moiety as AHFCA8 and could feasibly fit into the LBPs of MmfR/AvaL1, both the presence of a hydroxyl moiety in the alkyl chain region and the lack of a carboxylic acid moiety in the approximate position of the GBL ring likely results in the loss of binding activity through inability to form key interactions with the key tyrosine residues and subtle alteration in the hydrophobic region of the SCB1 molecule. In the case of MmA the similarity in the size of the MmA molecule in comparison with AHFCA2 similarly enables MmA to fit into the pocket of AvaL1, and through interactions *via* the carboxylic acid moiety to the key tyrosine residues exhibit minimal activity against AvaL1. Computational modelling of *in vivo* methylenomycin biosynthesis regulation performed by Dr Jack Bowyer[241] suggests that MmfR does not interact with MmA, corroborated from *in vivo* experiments with MmA and MmC (methylenomycin C; MmA biosynthetic precursor) performed by Dr Styles[149] (for interactions with both MmfR and MmyR). It is possible that the observed apparent activity of MmA against AvaL1 would not be capable of inducing the production of the azoxy compound *in vivo* (and by extension AHFCA2 given their similar activity), or that AvaL1 is more receptive to MmA than MmfR.

6.4.3 Inferring AHFCA specificity in other MmfR orthologues

SgnR and SclM1, when compared with the MmfR LBP residues HYL**A**F**M**AQPYWVFFQ, are identical in their LBPs save for two residues; HYL**V**F**V**AQPYWVFFQ in SgnR, and HYL**A**F**V**AQPYWVFFQ in SclM1. Assuming similar folding of helices to form the binding pocket, substitutions of these residues with similarly hydrophobic side-chained residues is not likely to result in a significant difference in AHFCA receptivity. The side-chain of the alanine-substituted residue contributes to the surface area of the alkyl-chain binding region, while the methionine residue, which is the most distal residue from AHFCA2 in the MmfR-bound crystal, appears to contribute little other than general hydrophobicity to the LBP. A substitution of the alanine residue with valine would further reduce the size of the cavity,

skewing recognition towards smaller AHFCAs. This is contrary to the AHFCAs produce in *S. venezuelae* (AHFCA5-8), which are closer in size to those produced by *S. avermitilis* rather than *S. coelicolor* A3(2). If SgnR responds to these AHFCAs, this would indicate that either:

a) Low affinity of the SgnR repressor for these larger AHFCAs is sufficient to fulfil its biological function, and therefore has thus not adapted specificity to these larger AHFCAs.
b) Substitution of the proximal residues surrounding the LBP leads to differences in folding resulting in an expansion of the LBP without significant mutation to the residues within it.

6.5 Conclusions

In corroboration with the earlier gel shift assays described in Chapter 3.3, the SPR experiments demonstrate not only that AvaL1 is an AHFCA receptor, but that it is most receptive with sub-μM IC_{50} values to the AHFCA signalling molecules AHFCA3, AHFCA5, AHFCA6 and AHFCA7 produced by the *S. lividans* TK24 pESAC13A-2H1 Δ*avaL2* heterologous host; the AHFCAs putatively produced by *S. avermitilis.* In contrast, AvaL1 was poorly receptive to AHFCA1, AHFCA2 and AHFCA4 produced by *S. coelicolor* A3(2), with AHFCA2 being the least active. The inhibitory strength of the four larger endogenous AHFCAs suggests that both AHFCA production and AHFCA receptivity are possibly co-evolved features, rationalised as AvaL1 mutations over time selecting for new AHFCAs produced by *S. avermitilis* due to changes in the specificity of the AvaA AHFCA biosynthetic enzyme for alternative β-ketothioester substrates.

In terms of AHFCA structure, AvaL1, like MmfR, was more strongly receptive to AHFCA signalling molecules with larger alkyl moieties, with preference for branched-isoalkyl molecules with alkyl-chains of the same length. AvaL1 however was considerably more sensitive to larger AHFCA molecules than MmfR, and smaller AHFCAs which produced stronger inhibitory responses against MmfR were diminished in activity against AvaL1. The differences in AHFCA response was rationalised *in silico* by an expansion of the hydrophobic channel of the AvaL1 LBP that surrounds the alkyl moiety of AHFCA effectors, resulting from substitution of several LBP residues, including an ArpA-subfamily conserved tryptophan residue.

The data also showed that AHFCA signalling molecules cooperatively inhibit the formation of the AvaL1:DNA complex. It is not possible to distinguish from this data the mechanism through which this occurs. Unfortunately, there is also not much data in the literature for the cooperativity of ligands binding to TFRs to compare against. The only example is of AcrR

directly binding ethidium, proflavine and rhodamine 6G, with n_H values of 1.02 ± 0.02, 1.12 ± 0.05 and 1.14 ± 0.09 respectively [219]. This direct binding to AcrR does not necessarily correlate with the ability of these effectors in inhibit AcrR DNA-binding, which may exhibit even greater cooperativity. Regardless, the data clearly shows that AHFCA molecules act cooperatively to inhibit the binding of DNA by AvaL1.

AHFCA4 showed similar inhibition of AvaL1 binding to DNA irrespective of the target ARE AvaL1 bound to, suggesting that AHFCAs may release AvaL1 and other orthologues from target AREs at similar rates *in situ*, with differences in observed transcription levels rationalised as a difference in promoter strength. All the tested signalling molecules showed similar behaviour across multiple chip-surface densities of immobilised DNA, indicating that inhibition by the signalling molecules is dependent only on the concentration of available AvaL1. The 6*R*-hydroxy-type GBL signalling molecule SCB1 had no effect on the DNA-binding ability of AvaL1, despite this GBL class sharing a biosynthetic route with AHFCA-type signalling molecules in *S. coelicolor* A3(2) and *S. venezuelae*[111, 135]. AvaL1 also responded poorly to the antibiotic MmA, consistent with MmfR.

Simulated interactions of AHFCAs 1-5 with MmfR modelled *in silico* with AutoDock Vina using the AHFCA2-bound MmfR structure had the strongest correlation with *in vivo* experimental data collected by Dr Styles, compared with alternate docking methods. Both the docking simulations and experimental data made it clear that the affinity of MmfR for AHFCAs gravitated towards an optimal length and branching of the AHFCA alkyl chain. The *in vivo* activity of AHFCA analogues (synthesised by Dr Malet) against MmfR strongly correlated with the docking simulations, with the exception of AHFCA analogues containing 4'-hydroxylmethyl modified moieties, which showed over-docking of ligand atoms within the proximity of H84. AHFCA specificity was modelled for MmfR and its orthologues using a wide range of *n*-alkyl, isoalkyl and *tert*-alkyl chained AHFCAs, including those both shorter and longer than AHFCAs known to be produced *in situ* by *Streptomyces* spp.

Modelling suggests SclM1 may have specificity for smaller AHFCA molecules, and SgnR may be more tolerant for larger AHFCAs (such as those produced by *S. venezuelae*), while strongly correlating with AHFCA docking scores for MmfR. By comparison, AHFCA docking to the AvaL1 homology model had no discernible correlation with the experimental data, as well as poor correlation with the docking data for other orthologues, indicating the inadequacy of the generated model because of distortions in the LBP resulting from disparate conservation of hydrophobic residues. Based on the strong correlations in docking scores and experimental data, structure-led design was used to generate hypothetical effector molecules

which would have stronger response against MmfR than the currently strongest inhibitor, AHFCA1. Four structures were generated, the best scoring of which was 4-isobutylfuran-2,3-dicarboxylic acid (cAJF8). Future work may elucidate a synthetic route for these molecules, which can then be tested for response against MmfR *in vitro*.

AHFCA4 did not stimulate or further inhibit the binding of AvaL2 to the DNA sites as has been previously suggested. MmyR and its orthologues lack many of the LBP residues found in MmfR orthologues, including the tyrosine, glutamate and ArpA-subfamily-conserved tryptophan residues (Figure 8-21, Appendix E), which strongly argues against the recognition of AHFCA molecules. The known binding of antibiotics to other pseudo-GBL receptors may indicate that antibiotics are the cognate ligands of these pseudo-AHFCA receptors[97]. The large size of the Phyre^2-modelled MmyR LBP compared with MmfR, which has a smallest predicted solvent-accessible area of 526 Å^2 (compared with 239 Å^2 for MmfR:AHFCA2), may be indicative of this (Figure 6-14). This putative antibiotic effector may either be exogenous, resulting in production of MmA as a response to possible threats, or another endogenous antibiotic as part of a coordinated network between biosynthetic gene clusters. Alternatively, there is the possibility that MmyR-type orthologues do not have an endogenous ligand, with these repressors serving only to control *in vivo* levels of their regulated secondary metabolites. Without a crystal structure it is not yet determined if there is a surface-accessible LBP.

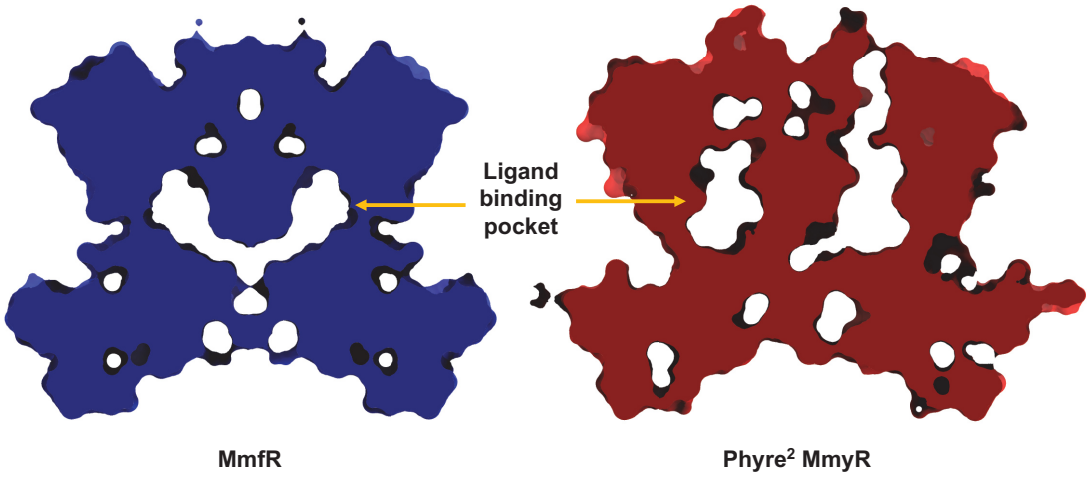


Figure 6-14 – Cross-sections of surface rendered structures of MmfR and the Phyre² MmfR-homology modelled MmyR. Each MmyR monomer was docked together using HADDOCK to form the homodimer.

7 Conclusions and future work

7.1 Final Conclusions

7.1.1 Biochemical characterisation of previously uncharacterised AHFCA and Pseudo-AHFCA receptors from *S. avermitilis* and *S. sclerotialus*

Reported in Chapter 3 is the *in vitro* preparation of the recombinant AHFCA receptors *S. avermitilis* AvaL1 and *S. sclerotialus* SclM1, and for the first time the preparation of a soluble recombinant pseudo-AHFCA receptor, *S. avermitilis* AvaL2 using pET151 cloning. AvaL1 and AvaL2 were readily soluble when expressed at temperatures between 15 °C and 37 °C, while SclM1 was only soluble when expressed overnight at low incubation temperatures (≤ 15 °C), as is also the case for the *S. coelicolor* A3(2) AHFCA receptor MmfR[139] and the *S. venezuelae* AHFCA receptor SgnR[148]. The pseudo-AHFCA receptor *S. sclerotialus* SclM4, like *S. coelicolor* A3(2) MmyR (Dr Rea, personal communication) was insoluble at all attempted expression temperatures, which makes AvaL2 currently the only pseudo-AHFCA available for use in *in vitro* studies.

The recombinant proteins were purified using gravity-flow IMAC, and AvaL1 and SclM1 binding to MmfR-like AREs within the AHFCA-regulated gene clusters of *S. avermitilis* and *S. sclerotialus* (identified using bioinformatics) was confirmed *via* gel shift assays (Chapter 3.3). SgnR was previously shown to bind AREs in both the *S. coelicolor* A3(2) *mmy* gene cluster and *S. venezuelae gbn* gene cluster[148], and here AvaL1 was shown to bind AREs in both the *S. avermitilis* azoxy compound gene cluster and the *S. sclerotialus scl* gene cluster, indicating a shared mechanism of DNA recognition.

Bioinformatic analysis (Chapter 5.2.2) indicates that AHFCA receptor orthologues across *S. coelicolor* A3(2), *S. venezuelae*, *S. avermitilis* and *S. sclerotialus* recognise and bind AREs with similar sequence compositions. The thirteen 24-bp palindromic AREs bound by AHFCA receptors have a consensus sequence of 5'-AAnATACCTTCG|CGAAGGTATnTT-3', and bioinformatics shows this consensus to be both similar and distinct from the consensus sequences of AREs bound by other ArpA-subfamily repressors (Chapter 5.2.3). Both the MmfR-type ARE and GBL/pseudo-GBL receptor ARE sequences share a similar consensus of A/T and C/G dinucleotide blocks (for GBL/pseudo-GBLS receptors a consensus sequence of 5'-WWWSSS|SSSWWW-3'; W = A/T and S = C/G), with the MmfR-type ARE motif deviating through the presence of a strongly-conserved A/T dinucleotide block occupying positions ±3 and ±4 near dyad centre (5'-AAnATACC**TT**CG|CG**AA**GGTATnTT-3'), resulting in a MmfR-type dinucleotide consensus sequence of 5'-WWWSWS|SWSWWW-3'.

Crystallisation trials performed with recombinant AvaL1, AvaL2 and SgnR (Chapter 3.4) yielded crystals of AvaL1 in the absence of any AHFCA ligands in buffer/salt-free crystallant conditions comprised only of Jeffamine® SD-2001 and 1-Propanol (MIDAS™ condition D12). However, x-ray diffraction of these crystals was not performed.

7.1.2 Decrypting control of AHFCA-regulated gene clusters in *S. avermitilis and S. sclerotialus*

As demonstrated by gel shift assays, three ARE sites were confirmed in the *S. avermitilis* azoxy compound gene cluster, and four AREs were confirmed in the *S. sclerotialus scl* gene cluster (Figure 7-1). This brings the total count of AHFCA-regulated gene clusters for which the AHFCA receptor AREs have been confirmed up to four; alongside three AREs within the *S. coelicolor* A3(2) *mmy* gene cluster[139] and three AREs within the *S. venezuelae gbn* gene cluster[148]. The *S. avermitilis* AHFCA receptor AvaL1 is proposed to regulate the azoxy compound gene cluster by inhibiting the transcription of genes within the AHFCA regulatory cassette (the transcriptional regulators *avaL1*/*avaL2* and the AHFCA biosynthetic enzymes *avaA/avaB/avaC*) as well as the gene cluster transcriptional activator *saverm_2301* in a manner analogous with that of *mmy* gene cluster regulation in *S. coelicolor* A3(2) by MmfR, which also includes a similar distribution of BldA-dependent TTA codons in the ORFs for the AHFCA biosynthetic enzymes and gene cluster transcriptional activator. In a manner similar to the regulation of the *S. venezuelae gbn* gene cluster by SgnR, the *S. sclerotialus scl* gene cluster is putatively regulated through transcriptional repression of all operons within the gene cluster by SclM1. Unlike the other AHFCA-regulated gene clusters no TTA codons are found within the genes of the *scl* gene cluster, so production of scleric acid may not be dependent on the accumulation of BldA.

Recombinant AvaL2 was unable to specifically recognise the same AREs bound by AvaL1 in both EMSA and SPR experiments (Chapter 3.3.4 and Chapter 4.2.1), indicating that AvaL2 likely regulates the azoxy compound cluster either through binding to other targets within the gene cluster or indirectly through a regulatory cascade. This is consistent with poor MmyR binding to *mmy* gene cluster AREs observed *in vivo*[149], but here the nonbinding of AvaL2 was more explicit. This challenges the initial hypothesis that the pseudo-AHFCA receptors recognise AREs bound by their AHFCA receptor counterparts. This is consistent with some GBL/Pseudo-GBL systems, such as the *S. coelicolor* A3(2) *cpk* GBL-regulated gene cluster, where the GBL receptor ScbR and pseudo-GBL receptor ScbR2 do not share all the same target AREs within the gene cluster[66, 76, 98, 99] (Chapter 1.3.5).

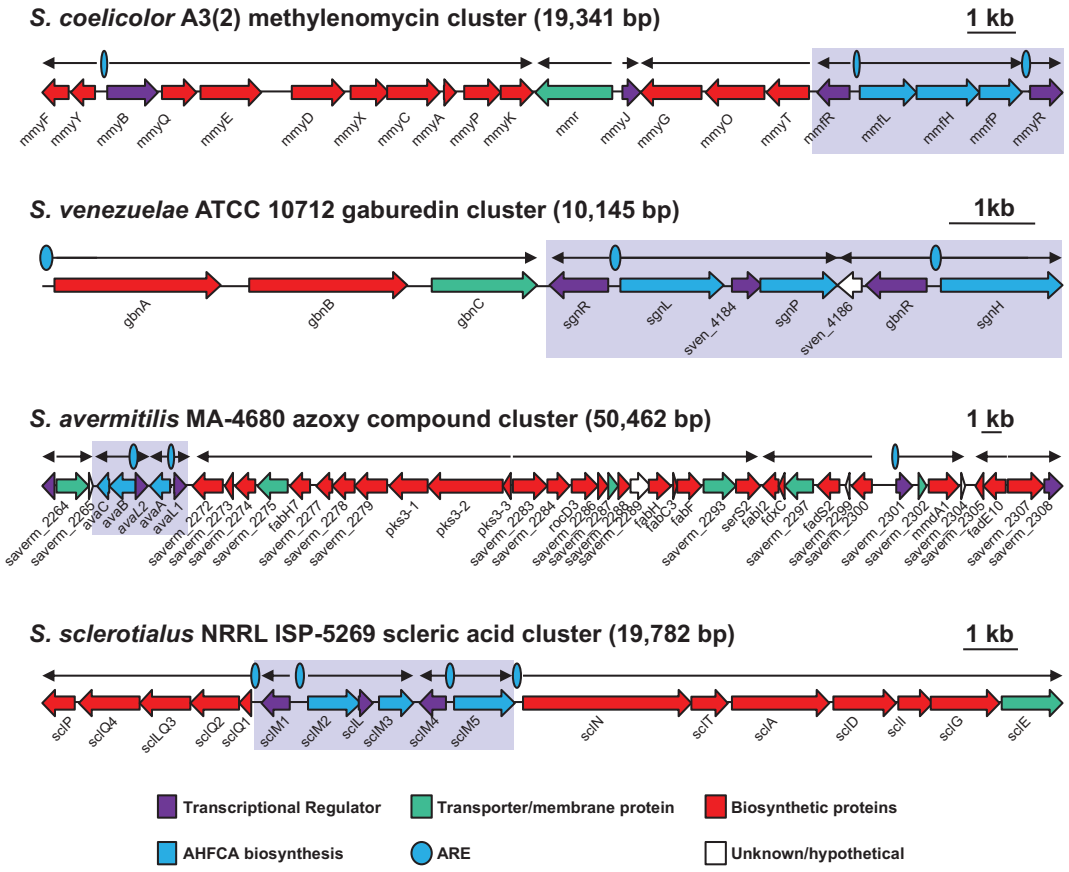


Figure 7-1 – Genetic organisation of the four *Streptomyces* AHFCA-regulated gene clusters annotated with confirmed AHFCA receptor AREs. Blue boxes indicate the AHFCA regulatory cassette, while black arrows indicate the putative transcribed operons.

7.1.3 Insight into the DNA-binding of an ArpA-subfamily repressor

In vitro quantification of an AHFCA receptor binding to target AREs is reported here for the first time (Chapter 4). Recombinant AvaL1 was shown to bind to all three AREs within the *S. avermitilis* azoxy compound gene cluster with a mean K_D value of 123.7 nM ± 11.2 nM at 25 °C, binding most strongly to the *avaL1/avaA*$_{ARE}$ with a K_D value of 110.6 nM ± 7.5 nM. This affinity is consistent with the reported affinity of the *S. griseus* GBL receptor ArpA binding to *ArpA*$_{CS}$ (apparent K_D of 110 nM to 215 nM)[204] and the *S. coelicolor* A3(2) pleiotropic GBL receptor CprB binding to *OPB* (150 nM) and *ArpA*$_{CS}$ (200 nM)[69]. The binding of SclM1 to the AREs within the *S. sclerotialus scl* gene cluster measured semiquantitatively *via* gel shift assays also suggest an affinity in the low hundreds of nM (Chapter 3.3.8), similar to that of AvaL1. It is anticipated that many other uncharacterised members of the ArpA-subfamily share similar binding affinities with their cognate ARE targets, including the pseudo-AHFCA receptors for which a target has not yet been confirmed.

AvaL1 bound to these AREs with a mean stoichiometry of 4.04 ± 0.62 AvaL1 monomers per ARE bound, indicating that AvaL1, like CprB, binds to DNA in a configuration consistent with a homodimeric pair of dimers[69, 152]. Early CryoEM images have indicated that MmfR similarly binds to DNA as a pair of homodimers (Dr Bhukya, personal communication), and the gel shift assays suggest that SclM1 also binds to DNA as a pair of homodimers (requiring further quantitative confirmation). This would appear to indicate that all AHFCA receptors bind their target AREs as a pair of homodimers.

Kinetic analysis shows that AvaL1 does not interact with DNA *via* a 1:1 Langmuir mechanism, but instead *via* a biphasic two state mechanism where the binding of AvaL1 homodimers to DNA alters the equilibrium between free AvaL1 and AvaL1 bound to the target ARE. The two state model accounts for this through a rate-limiting conformational shift which switches the AvaL1:ARE complex between an unstable and stable AvaL1:ARE complex. The orders of magnitude for the determined rate constants were $k_{a1} = 10^5$ M^{-1} s^{-1} and $k_{d1} = 10^{-2}$ s^{-1} for the initial binding, and for the conformational shift step $k_{a2} = 10^{-3}$ s^{-1} and $k_{d2} = 10^{-3}$ s^{-1}. The kinetics for the binding event were recently utilised in a paper published by Bowyer *et al*[241] which computationally modelled the methylenomycin regulatory system in *S. coelicolor* A3(2). The modelled rate constants for MmfR binding to its target AREs was consistent with the experimental data for the production of methylenomycin *in vivo*, which substantiates the kinetics reported here.

This conformational shift component within the kinetic model is consistent with the clamp-and-click model of TFR binding[172], whereby upon binding of each TFR dimer to the target operator the TFR dimers undergo interdimeric twisting resulting in local undertwisting of DNA, facilitating docking of the α3 recognition helices into the DNA major grooves[172]. However, a review of SPR sensorgrams for TFRs binding to DNA in the literature suggests that this biphasic binding is only observed in TFRs which bind to DNA as a pair of dimers. Therefore, this altered equilibrium between free TFR and complexed TFR is likely dependent on the bound stoichiometry of TFR dimers, for which the described in clamp-and-click model still applies but may be more complex than previously indicated.

AvaL1 was shown to bind to DNA across a temperature range of 7 °C to 40 °C, with thermodynamic analysis of the binding kinetics indicating that the binding reaction was exothermic and driven by both enthalpy and entropy ($-\Delta H$, $+\Delta S$), whereas the conformational shift was endothermic and driven by entropy ($+\Delta H$, $+\Delta S$). It is proposed that the initial binding event provides the necessary energy to drive forward the conformational shift, with the net reaction being exothermic and driven by enthalpy ($-\Delta H$, $-\Delta S$). ITC analysis for CprB binding

to *ArpA*$_{CS}$ and *OPB* showed two different modes of binding, with CprB binding to *ArpA*$_{CS}$ fitted using a single site model, and binding to *OPB* using a two site model[69]. For CprB binding to *ArpA*$_{CS}$ the net interaction was also exothermic and driven by enthalpy (-Δ*H,* -Δ*S*), and for CprB binding to *OPB* the first site complex formation was exothermic and driven by enthalpy (-Δ*H*, -Δ*S*), while the second site complex formation was endothermic and driven by entropy (+Δ*H*, +Δ*S*). This is almost analogous with the thermodynamics derived for AvaL1 binding to DNA, but it is not certain how comparable the two models really are.

It was noted that these different modes of DNA-binding for CprB correlated with sequence palindromacy, with positive cooperative binding of CprB to *OPB* ($n_H \approx 1.6$) observed due to the poor palindromacy of the *OPB* sequence, compared with binding to *ArpA*$_{CS}$ which is highly palindromic. AvaL1 binding to highly palindromic AREs exhibited curvilinear cooperativity (correlating with the availability of DNA) consistent with the *M. tuberculosis* TetR-family repressor EthR[206], with an $n_H \approx 1$ at [AvaL1] = K_D. A hypothetical model was proposed whereby cooperativity of ArpA-subfamily repressors binding to DNA is observed as a function of the palindromacy of the target sequence, with ArpA-subfamily repressors binding to highly palindromic sequences noncooperatively through simultaneous interactions of each dimer, and binding to poorly palindromic sequences through necessary positive cooperativity to enable sequential/simultaneous binding of the first dimer to the motif-matching half-dyad which better allows the second dimer to recognise the poorly recognised half-dyad.

MmfR HADDOCK docking to the AHFCA receptor consensus sequence 5'-AATATACCTTCG|CGAAGGTATATT-3' (Chapter 5.3) resulted in localisation of the MmfR homodimer to the DNA region occupied by CprB dimer A in the CprB:*ArpA*$_{CS}$ crystal structure[69]. Both CprB and the MmfR model utilised many of the same amino acid residues (in relative alignment) to form the majority of electrostatic interactions with the phosphate backbone through the spacer helix (α2) and helix α3. The MmfR chain A helix α3, docked into the major groove in the region of the highly conserved C nucleotide at position -6 (5'-AATATA**C**CTTCG|CGAAGGTATATT-3'), with residue Y67 forming electrostatic interactions with both C-6 and its complementary nucleotide. HADDOCK was unable to simulate the dimeric twisting necessary to fit the MmfR chain B α3 helix into the major groove in the other half-dyad, instead docking over the adjacent minor groove. This would support the clamp-and-click model, whereby anchoring to the DNA phosphate backbone *via* interactions with helix α2 is proposed to facilitate the twisting of the dimeric interface, which in turn undertwists the major groove to enable fitting of the chain B α3 helix.

If MmfR homodimers interact with palindromic DNA simultaneously the initial binding of dimers might involve anchoring to the regions surrounding the highly conserved C-6 and G+6 nucleotides by the chain A α3 of each dimer, which may either independently or interdependently induce local undertwisting in the distal major groove to facilitate the docking of the chain B α3 helix. In sequential binding the first dimer may bind as suggested for a single dimer, with the subsequent undertwisting of DNA around the chain B α3 helix enabling the fitting of the chain A α3 helix for the second dimer within that region.

From this docking and sequence homology of the DBD with other TFRs, predictions of DNA-interacting residues were made for other ArpA-subfamily repressors, with the assumption that these repressors also share an underlying binding mechanism. Noted also from the available crystal structures of other ArpA-subfamily regulators such as CprB and TylP[(178)] was that other factors may influence specificity of sequence recognition, including the distribution of electrostatic potential across the DBD interface and the geometry of the DBD.

7.1.4 AHFCA receptors are specifically receptive to endogenous AHFCA autoregulators

The assignment of SgnR, AvaL1 and SclM1 as AHFCA receptors was initially made based on protein sequence homology with the LBD of MmfR. The inhibitory activity of AHFCAs against SgnR was previously confirmed *in vitro* by Dr Zhou[(148)], and here both EMSA (Chapter 3.3.7) and SPR (Chapter 6.2) have demonstrated the inhibitory activity of AHFCA signalling molecules against AvaL1. Determination of AHFCA receptivity for the SclM1 receptor was inconclusive but given the strong sequence homology of the SclM1 LBP with the LBPs of MmfR and SgnR it is likely that SclM1 also functions as an AHFCA receptor.

SPR was used to measure the inhibition of AvaL1 DNA-binding by AHFCA(1-5) produced by *S. coelicolor* A3(2)[(95)] and AHFCA(3, 5-7) produced by *S. avermitilis*[(112)]. AvaL1 binding to DNA was most strongly inhibited in the presence of the endogenous *S. avermitilis* AHFCA autoregulators (mean IC_{50} = 0.48 μM ± 0.06 μM), and less inhibited by the uniquely produced *S. coelicolor* A3(2) AHFCAs (mean IC_{50} = 2.06 μM ± 0.14 μM for AHFCA(1, 4) and 18.98 μM ± 3.38 μM for AHFCA2). In addition, some non-specific inhibition was observed by the addition of the antibiotic MmA (IC_{50} = 25.18 μM ± 16.62 μM), and AvaL1 binding was nonresponsive to the 6*R*-hydroxy-type GBL SCB1 produced by *S. coelicolor* A3(2). DNA-binding by MmfR was similarly unaffected by SCB1[(149)], suggesting that coexisting GBL and AHFCA receptors respond only to their specific effectors (although there is no confirmation of GBL receptors being nonreceptive to AHFCA signalling molecules). Inhibition by

AHFCAs was also cooperative for all AHFCA autoregulators regardless of source organism (more so in the endogenous *S. avermitilis* AHFCAs; mean $n_H = 1.77 \pm 0.35$), and the observed inhibition was independent of both the saturation of DNA by AvaL1 and the ARE sequence being bound to.

Through the analysis of AHFCA SARs in corroboration with both *in silico* AHFCA docking studies and experimental data presented both here and by previous Corre/Challis group members (Dr Styles[149], Dr Zhou[148] and Dr Malet[139]), it was determined that there is an optimal configuration of AHFCA alkyl chain length and methyl branching for activity against both MmfR and AvaL1, with AvaL1 being more receptive to the larger AHFCAs produced by *S. avermitilis*, and MmfR the smaller AHFCAs produced by *S. coelicolor* A3(2). This data correlated with deviations in the sequence conservation of the AvaL1 LBP compared with the MmfR LBP, whereby residue substitutions in the AvaL1 LBP resulted in an expansion of the LBP in the region occupied by the AHFCA alkyl moiety. The receptivity of the AHFCA receptors for different AHFCA structures is possibly co-evolved with either differences in the AHFCA biosynthetic machinery or in *Streptomyces* spp. fatty acid metabolism; for which future work will be required to elaborate upon. Each *Streptomyces* sp. likely produces the optimum AHFCA structure for their specific AHFCA receptor, which for SgnR will be one of the AHFCAs amongst AHFCA(5-10) (excluding AHFCA8E) which are produced by *S. venezuelae*[111].

7.1.5 Deconvoluting current models of the AHFCA regulatory cascade

The new *in vitro* data suggests that previous assumptions made concerning the differences in the expression levels of targets regulated by AHFCA and pseudo-AHFCA receptors in prior *in vivo* experiments may be incorrect. While there is now corroborating *in vivo* and *in vitro* evidence to suggest that the AHFCA receptors bind to the ARE upstream of the AHFCA receptor gene the strongest, the difference in AvaL1 binding between the three characterised *S. avermitilis* AREs was not significant enough to account for the observed difference in previously observed *in vivo* activity[67, 111, 149]. *In vivo* experiments performed by Dr Styles suggested also that AHFCA receptors bound at different AREs required different AHFCA amounts to be released from DNA, but the *in vitro* data here suggests that the ARE bound by the AHFCA receptor is independent of AHFCA activity. Parallel *in vivo* experiments indicate that the expression levels of AHFCA receptor targets are influenced not by the binding of the repressor to these AREs, but rather the strength of the promoters to which they block RNAP from binding to (Dr Rodríguez García, personal communication).

In addition, the role of the pseudo-AHFCA receptor within these gene clusters still remains ambiguous. The lack of observed *in vitro* binding of AvaL2 to the AHFCA receptor AvaL1 AREs does corroborate with the lack of *in vivo* repression of *lux* genes by MmyR as reported by Dr Styles, but AvaL2 did not show any change (positive or negative) in DNA-binding response in the presence of AHFCAs (assayed with AHFCA4) as proposed for MmyR by Dr Styles[149] and GbnR by Dr Sidda[111]. This suggests that any change in the *in vivo* output of target genes were not likely due to direct interactions of pseudo-AHFCA receptors with AHFCAs, but through secondary effects; whether that be metabolic or regulatory in nature.

Bioinformatics analysis using MEME-Suite was performed to predict possible targets within the genomes of *S. avermitilis* and *S. coelicolor* A3(2) (Chapter 5.2.6), including predictions of other AREs within AHFCA-regulated gene clusters and within regions bound by other ArpA-subfamily repressors (to investigate cross-talk of ArpA-regulated systems). While many ARE predictions were made, none of these were as statistically significant as the AREs within the AHFCA-regulated gene clusters confirmed to be bound by the AHFCA receptors. In addition, many of the sequences predicted featured half-dyads matching at least one half-dyad of the MmfR-type motif; a feature observed in sequences bound also by ScbR and ScbR2 (Chapter 5.2.4)[76]. Given also the dissimilarity in match for each ArpA-subfamily repressor ARE motif between gene cluster-specific and pleiotropic targets (Chapter 5.2.3), it is possible that some sites cannot be predicted using bioinformatics alone, so may need corroborating transcriptional studies to validate and/or identify dissimilar but recognised ARE sequences. Attempts were made to validate these predictions using SPR to screen a refined library of *S. avermitilis* AREs (as previously predicted by Dr Poon[112]), but the results were inconclusive; two possible sites were identified for AvaL2, but further validation is required. As such, the putative role of both AHFCA and pseudo-AHFCA receptors as pleiotropic regulators of secondary metabolite biosynthesis in morphological differentiation in *Streptomyces* spp. remains undetermined.

7.1.6 Applicability of SPR-based methodologies in studying TFRs

In this thesis both analytical and experimental methodologies have been established which allows the prediction of important regulatory sites within AHFCA-regulated gene clusters of interest using bioinformatics, and both the confirmation and quantification of these interactions using SPR. SPR has proven to be a versatile methodology for not just quantifying the affinity of direct interactions between AHFCA receptors and their target AREs, but also in characterising affinity, kinetics, thermodynamics, stoichiometry, cooperativity and binding mechanisms using SA-immobilised dsDNA hairpin oligonucleotides. Interactions between

repressors and DNA could also be used to investigate the activity of the AHFCA signalling molecules through inhibition of repressors binding to DNA. All of the above listed characteristics could also be assessed for the direct interaction between AHFCAs and their target AHFCA receptor utilising NTA-chip immobilised protein, but further optimisation of the SPR system through stabilisation of receptor immobilisation (through incorporation of a longer polyhistidine-tag such as His_{10} or His_{12}) on the NTA-chip surface is required.

Immobilisation of protein could also be used for high-throughput screening of binding between ArpA-subfamily repressors and putative ARE sequences without the need for labelling of oligonucleotides (*e.g.* biotin). While the results were inconclusive, the methodology remained viable. This could be a powerful tool for the future study of not just other ArpA-subfamily repressors, but also of orphan TFRs for which both the DNA-binding and effectors are not known. As many TFRs are autoregulatory, a high number of binding sites can be discovered by screening of the promoter regions for these TFR genes, and from these other targets can be predicted using bioinformatics and then rapidly screened. Effectors can then be screened for TFRs for which the binding sites are known, including other members of the ArpA-subfamily.

However, it is clear that SPR is not a solve all solution for all possible research questions. While SPR is advantageous in the analysis it provides as well as the small sample sizes required, one of the biggest drawbacks of SPR is the combined cost of the instrument, sensors, maintenance, and technical expertise required to maintain and operate the system. In addition, the interpretation of the SPR is somewhat limited by the lack of SPR data available for the analysis of DNA-binding by other ArpA-subfamily repressors with which to directly compare against, for which the most conclusive data (for CprB) has been collected by ITC[(69)]. In addition, SPR has some limitations in characterising the mechanisms of interactions involving multivalent interaction partners. It is hoped in the future that more data utilising both SPR/ITC and other methodologies can be used to corroborate and deconvolute the findings in this thesis.

7.2 Proposed future work

7.2.1 Exploiting the regulatory AHFCA cassette in biotechnology

One of the research focuses of the Corre group is discerning ways that AHFCA regulatory cassettes can be exploited in synthetic biology & biotechnology. The concept of exploiting GBL and other quorum sensing signalling systems for use in synthetic signalling circuits has seen increasing interest in recent years[(242)]. ScbR has already been used for non-toxic

transcription induction in mammalian cells *via* fusion of ScbR with eukaryotic transactivator domain VP16[(243)], complex systems utilising multipartite ScbR/TetR/TtgR/VP16 chimeras[(244)], and construction of synthetic orthogonal logic gates[(200)]. Other possible applications for GBL systems in synthetic circuits include the use of multiple GBL receptors, each recognising different GBLs (*i.e.,* one for 6-keto type GBLs, one for 6*R*-hydroxy type GBLs, *etc*). All the same considerations for GBL regulatory systems are also applicable to AHFCA regulatory systems, which could also be used in synthetic circuits alongside GBL receptors. One of the key advantages of both AHFCA and GBL regulatory systems is that only a few genetic components are required to produce the signalling molecules, the biosynthetic genes for which are conveniently colocalised with each other and their target receptors within the regulatory cassette, which can not only be exploited in synthetic systems also as a source of AHFCAs/GBLs for applications in research and biotechnology.

Exploitation of the AHFCA regulatory cassette as a novel protein expression system is currently being investigated by the Corre group. Because of the autoregulatory nature of AHFCA/GBL cassettes, once the minimal induction concentration threshold is crossed the derepression of AHFCA/GBL biosynthetic genes results in the production of further AHFCAs/GBLs, thus self-maintaining the expression of both AHFCA/GBL biosynthetic genes as well as other target genes. Expanding upon this, these AHFCA/GBL regulatory cassettes could also be used to develop an autoinducible protein expression system. Autoinduction removes the need for carefully-timed inoculation of bacterial cultures with the inducer by instead either enabling the production of the autoregulator by the bacteria itself or through addition of the inducer into the expression media[(245)]. Protocols for autoinduction in T7 expression systems using pET vectors in *E. coli* BL21(DE3) have previously been described by Studier[(246)], and other autoinduction systems which have been developed for protein expression in *E. coli* include the lactose-driven SILEX system (modulated by glucose)[(247)], as well as the pLAI system which utilises the quorum sensing signalling molecule N-3-(oxohexanoyl)-l-homoserine lactone as an autoregulator[(248)].

AHFCA/GBL-based systems could also be used for protein expression in Gram-positive hosts such as *Streptomyces*, which would enable expression of proteins not easily expressed in Gram-negative expression hosts like *E. coli* (a result of multiple factors; *e.g.*, rare codons, lack of appropriate chaperones, different cytoplasmic compositions, *etc.*), such as the *Streptomyces* pseudo-AHFCA receptors. The AHFCA expression system could also be exploited for engineering biosynthetic pathways to maximize biosynthetic efficiency by refactoring the regulatory components of gene clusters of interest, and the utilisation of BldA by *Streptomyces* could be exploited for tighter time-dependent expression *via* incorporation of TTA codons.

However, possible undesirable effects through pleiotropic regulatory activity by the AHFCA/pseudo-AHFCA may be created; hence the necessity for investigating pleiotropic regulation of these repressors.

The AvaL1/AvaL2 repressors may be the strongest candidates for expression systems in an *E. coli* expression host, based on the solubility of the proteins as well as stable of AvaL1 binding activity binding over a 7 °C to 40 °C temperature range. If the activity of AHFCAs against MmfR determined *in vivo* by Dr Styles is comparable to its *in vitro* affinity[149], then AvaL1 may be a better choice due to its apparent increased receptivity to AHFCAs produced by *S. avermitilis.* However, this would need to be balanced against the metabolic cost of producing said AHFCAs. Because of cross-AHFCA receptivity, the biosynthetic machinery from different systems can be easily exploited to select for the production of specific AHFCA molecules suitable for the expression host.

Another possible avenue of research considered was the design of novel inducers based on both *in silico* studies and SAR studies performed in this project. However, the cost and time associated with both developing a route through organic synthesis and then measuring their activity greatly limits their viability, especially considering that that the biosynthetic machinery for already very strong inducers is already available, which may or may not be capable of producing these novel inducers. It is possible that the AHFCA biosynthetic proteins could in the future work be manipulated to generated novel AHFCA-like compounds, first by determining the selectivity of AHFCA precursor substrates and then direct mutagenesis of the enzymes. However, time and resources may be better spent optimising and improving the selectivity and activity of AHFCAs already produced by the AHFCA biosynthetic machinery through mutagenesis of the AHFCA receptor itself.

7.2.2 Biophysical and structural studies

Although the biological activity of the expressed recombinant AvaL1 was established, the amino acid sequence of the expressed recombinant protein would ideally have been confirmed using mass spectrometry (to corroborate the GATC sequence confirmation during cloning). Analytical ultracentrifugation (AUC), size-exclusion chromatography (SEC) or SEC-MALS (multi-angle light scattering) could have been used to confirm both the molecular weight of the produced protein as well as investigate oligomerisation of TFR dimers. Buffer optimisation for protein solubility/DNA-binding was not performed for the DNA-binding experiments in these studies; EMSA and SPR buffers containing HEPES were selected for measurement of DNA-binding based on general buffers used in SPR as well as the previous

use of HEPES buffer in gel shift assays for both MmfR and SgnR[148]. We were satisfied that with AvaL1 maintaining strong DNA binding ability (compared to published binding affinities for other ArpA-subfamily repressors and previous gel shift assays) after a series of buffer exchanges (KH_2PO_4/K_2HPO_4 to Tris-HCl to HEPES) in gel shift assays performed in both HEPES (as observed here and in previous work) and Tris-HCl (gel shift assays in undergraduate labs and as performed by previous Corre/Challis group members[139]) binding buffers, that this would serve as an indicator of adequate protein stability. For confirmation of this however buffer optimisation would have been performed in hindsight, and in future should be performed for both AvaL1 and other ArpA-subfamily repressors; especially those less soluble when expressed. Thermal shift assays could be performed to explore the stability of biological activity for TFRs under different temperature and buffer conditions, as could SPR given the use of appropriate blank and binding buffers.

In hindsight it would also have been useful to use SEC to measure aggregation of proteins in solution. Circular dichroism (CD) could have been used to measure effects such as thermal, freeze-thawing, or pH stability, using the theoretical CD spectra generated from the MmfR pdb structure file using a tool such as DichroCalc[249] (Figure 8-22, Appendix F) for reference. CD could also have some for assessing structural conformations assumed by the AHFCA receptors in complex with both DNA and effectors, which could also be used to study the effects of mutagenesis on protein stability and activity.

The crystal structure of the AvaL1 repressor can be determined now that crystallisation conditions are known. Given the homology of AvaL1 with MmfR (42.5 % sequence identity) it is feasible that any diffraction patterns collected from the crystals of AvaL1 could be resolved using molecular replacement. Given the ability of MmfR to cocrystallise with AHFCA2 in identical crystallant conditions, AvaL1 may exhibit similar behaviour in cocrystallisation trials with other AHFCA signalling molecules. This would corroborate the SAR studies here for AHFCA receptivity. AvaL1 bound with different AHFCA molecules may enable determination of whether the degree of shift between the ligand-free and AHFCA-bound configurations is different for each AHFCA molecule, or dependent solely on AHFCA binding within the LBP.

Crystallisation trials of an AHFCA receptor in complex with a DNA target should also be considered to help us better understand the mechanisms of DNA-binding. AvaL1 has been shown to bind to a fully palindromic consensus sequence (Chapter 3.3.3), which is more suitable for crystallography due to better packing within the crystal matrix. The acquisition of an AvaL1 or MmfR crystal structure in complex with DNA would help verify the model of

DNA-binding as proposed in Chapter 5.3. Nuclear magnetic resonance (NMR)-derived structures could be used to explore the range of proteins dynamics for MmfR/AvaL1 in unbound, ligand-bound and DNA-bound states. Using structures derived from X-ray crystallography or NMR could also be utilised in molecular dynamics simulations to deconvolute binding mechanisms, such has been used for CprB binding to DNA, such as has already been used in such a manner to investigate the involvement of the positively charged N-terminal region of the CprB DBD in stabilising the binding of CprB to *OPB*[152]. AUC, SEC or SEC-MALS could also have been used to determine whether binding of dimers to DNA was sequential or simultaneous, as would be observed by formation of individual complexes. As previously mentioned, Cryo-EM has been used to capture images of MmfR in complex with DNA and could also be used to verify stoichiometric configurations of AHFCA/pseudo-AHFCA receptors in complex with DNA.

Crystallographic trials have not yet yielded crystals for AvaL2, but precipitation of protein has been observed for some crystallant conditions, so crystallisation of AvaL2 remains a future prospect. Despite being of different receptor functions, AvaL2 still retains a 30 % sequence identity with MmfR, so there may still be hope for molecular replacement to solve the phase problem for AvaL2 diffraction patterns. Any crystal structures obtain for the repressor can be further utilised *in silico* to identify the cognate effectors of this regulator. Crystallisation trials of AvaL2 in complex with DNA can be attempted when future binding sites are confirmed.

7.2.3 Further *in vitro* characterisation of AHFCA/Pseudo-AHFCA receptor interactions with DNA

SPR could be used to compare DNA-binding affinity, kinetics, stoichiometry, thermodynamics, cooperativity and mechanism for each AHFCA receptor that has thus far been expressed. Apart from CprB, there is little quantitative data for the DNA-binding of ArpA-subfamily repressors to either their target AREs or effectors. SPR would be a good tool to quantify the interactions of other ArpA-subfamily repressors which are soluble *in vitro*. It is proposed that ITC experiments also be performed for AHFCA repressors, for which a comparison can be made with the ITC data collected for CprB and hopefully help to deconvolute the binding mechanisms of ArpA-subfamily repressors. It may also be a point of interest to compare cross-binding of AREs bound by different repressors from the ArpA-subfamily.

With recombinant AvaL2 current being the only pseudo-AHFCA receptor to be solubilised, it is the only pseudo-AHFCA receptor for which DNA-binding can be studied *in vitro*. As such,

this would make it representative of other pseudo-AHFCA receptors if used for *in vitro* studies. The DBDs of the pseudo-AHFCA receptors from the four AHFCA-regulated systems thus far share approximately 60 % to 70 % sequence identity between them, and the recognition helix is fully conserved between all of them, so inferring DNA recognition of other orthologues based on *in vitro* AvaL2 binding experiments may not be not completely unreasonable. However, the sequence identity of the LBDs of pseudo-AHFCA receptors range from approximately 20 % to 40 %, with SclM4 and GbnR having the most similarity and GbnR/MmyR the least. Therefore, there is no guarantee that AvaL2 would be a suitable substitute for the characterising the effector affinity and specificity of other pseudo-AHFCA receptors. Once a target sequence is confirmed, SPR can be used to characterise this repressor. Two putative targets have been identified in this project, confirmation of which would be a strong starting point for investigating the AvaL2 repressor.

Investigating the binding of AHFCA/Pseudo-AHFCA receptors to AREs with half-dyad sequences raises some interesting questions and implications, such as the following:

- How can we refine/expand ARE predictions/identification using bioinformatics?
- Could predicted half-dyad AREs in the AHFCA-regulated gene clusters be targets of these repressors?
- How does DNA-binding affinity for AHFCA/pseudo-AHFCA receptors change with ARE palindromacy?
- How would differences in sequence palindromacy reflect on the proposed model for dimer cooperativity?
- Would binding stoichiometry be disrupted, as for RamR?[211]

A similar approach to that used by Onaka and Horinouchi to study the interactions of *S. griseus* ArpA with DNA could be investigated[204], whereby a library of oligonucleotides is screened/captured, with each oligonucleotide sequence containing one half-dyad with a fixed nucleotide sequence, and another half-dyad which is randomised, rearranged, inverted, *etc*. Predicted half-ARE sequences within AHFCA-regulated gene clusters would also be good screening targets. A number of methods could be used to test not just half-dyad dependency but also a wide range of ARE sequence composition experiments; microarray-based technologies such as ChIP-on-chip have been previously proposed for this purpose, but a wide range of other methods could also be applied, including SELEX (systematic evolution of ligands by exponential enrichment), CASTing (cyclic amplification and selection of targets), Bind-n-Seq, PBM (protein binding microarray), CSI (cognate site identifier), MITOMI (mechanically induced trapping of molecular interactions), and others[250].

Other methods of measuring the binding properties of these repressors could be used to provide alternative assessments that may corroborate or add to the analyses shown here. For thermodynamics analysis the application of ITC for DNA binding has already been discussed, especially as a comparative with CprB binding to DNA[69], which can be used to deconvolute the models of DNA binding model between AvaL1 and CprB. WISB had also recently acquired a Monolith NT.115 for measuring binding via Microscale thermophoresis (MST), which could also be used for investigating affinity and thermodynamics.

7.2.4 Investigating heterodimeric configurations

It is possible that AHFCA/pseudo-AHFCA repressors, like ScbR/ScbR2[156] and *S. avermitilis* AvaR1/AvaR2[78], form heterodimeric complexes with one another. Oligomerisation equilibrium studies using AUC to determine concentration-dependent monomer dimer stoichiometry and affinity for heterodimeric assemblies are potential future experiments. Because the repressors are of similar *mW* however it may not be practical to try and determine dimerisation of heterodimers for native proteins with the current pET151 constructs, but a methodology could be applied based on the competitive DNA-binding assays performed for MmfR DNA-binding by Dr Zhou[148], whereby new protein constructs could be generated utilising high *mW* tags to generate recombinant protein monomers of different *mW*. GST pulldown in combination with western blot analysis, as done for AvaR1/AvaR2, is another option[78]. SPR could potentially be used, but the methodology would have to be assessed before meaningful data could be collected.

Because of an assumed equilibrium between AHFCA/pseudo-AHFCA repressors in both homodimeric and heterodimeric configurations, measuring DNA-binding quantitatively may be difficult. Therefore, semiquantitative methods like EMSA may be better employed to characterise heterodimeric/homodimeric complex formation. If heterodimerisation results in the binding of novel ARE sequences (as was observed for the ScbR:ScbR2 heterodimer[156]), such as AREs with half-dyad sequences, this would have further implications in bioinformatic prediction of AREs as well as the characterisation of both regulatory models and DNA-binding mechanisms. It would also interesting if ArpA-subfamily repressors could heterodimerise with not just their partner repressors, but with other ArpA-subfamily repressors as well. For example, given that six ArpA-subfamily repressors are known to be encoded in *S. coelicolor* A3(2), there are a theoretical 21 combinations comprised of 6 homodimers and 15 heterodimers, all of which may bind potentially novel AREs, and would also potentially exhibit bipartite effector receptivity.

7.2.5 Identifying and characterising AHFCA/Pseudo-AHFCA receptor effectors

The AHFCA receptivity of MmfR, SgnR and SclM1 could be characterised *in vitro* using SPR as described in this thesis. This would include measuring the inhibitory efficacy of AHFCAs from each *Streptomyces* sp. against each repressor to test the hypothesis of co-evolved AHFCA receptivity/production. This would also be an opportunity to further quantify the interactions of synthetic AHFCA analogues, such as those designed *in silico* or those synthesised by Dr Malet[139]. The direct measurement of AHFCA receptor:AHFCA interactions could also prove insightful in better understanding the mechanism of binding.

The effector of AvaL2 and other pseudo-AHFCA receptors still remains elusive. The current hypothesis based on functional homology with pseudo-GBL receptors is that AvaL2 and other pseudo-AHFCA receptors act as antibiotic receptors, binding one or more antibiotics which may be endogenous, exogenous or both. Two methods are proposed for identifying these effectors; co-purification and SPR. Co-purification has already been used to capture SCB-type GBLs which are bound by ScbR[251, 252]. Effector activity could be screened by SPR using an SA-chip based methodology with DNA immobilised to the chip, provided a target sequence can be identified, or the effector could be captured using an NTA-based methodology with AvaL2 immobilised to the chip, then regenerated from the chip surface and the fraction recovered. Crude *S. avermitilis* organic extracts (both induced with AHFCA and uninduced) could be screened to determine if the effector is endogenous and depending on *S. avermitilis* culture conditions could be used to determine how AvaL2 activity is coordinated with other cellular processes. Screening could also be performed with a library of known antibiotic compounds, starting with known effectors of other pseudo-GBL receptors such as jadomycin B, chloramphenicol, actinorhodin, undecylprodigiosin, *etc*[97]. Assuming the effector is exogenous, it is likely an antibiotic or signalling molecule produced by another bacterium. A method based on screening cell-culture extracts could be used to co-culture *S. avermitilis* with a bacterial species that is able to stimulate the production of the azoxy compound in *S. avermitilis*[253].

7.2.6 Investigating pleiotropic regulatory roles of AHFCA/pseudo-AHFCA receptors

While it has been demonstrated that the AvaL2 repressor does not bind to the same sites as AvaL1, it is still unclear how exactly the AvaL2 repressor comes to control the azoxy compound gene cluster in tandem with AvaL1, or how, if it does so, coordinate gene cluster regulation with other secondary metabolite biosynthetic pathways. It is still hypothesised that

AHFCA/pseudo-AHFCA receptors function as pleiotropic regulators, based on the observed pleiotropic regulatory activity of other TFRs (Chapter 5.2.4). Despite the apparent lack of binding observed for pseudo-AHFCA receptors to AHFCA receptor AREs, these pseudo-AHFCA receptors are clearly important in the negative regulatory feedback of their gene clusters. Analysis of AHFCA/pseudo-AHFCA receptor deletions in *Streptomyces* spp. using RNA-seq based transcriptomics may reveal many of the target genes regulated by these repressors both within the AHFCA-regulated gene cluster and at other regions within the genome, which can be later be identified using bioinformatics and then confirmed *in vitro*. However, as discussed in this thesis, competition for binding sites with other regulators (including other ArpA-subfamily repressors) may interfere with identification of these targets. A bioinformatics-led approach using ChIP-on-chip or ChIP-seq technologies (or other microarray based methods as mentioned prior) could be used for high throughput screening of predicted AREs.

SPR could also be used to screen binding to bioinformatically-predicted AREs but given the number of predictions using SPR this may be impractical. Using SPR for screening sites, bioinformatically-predicted or otherwise, needs to either needs to have the number of predictions bioinformatically refined (as discussed in Chapter 5.2.6), or be complimented with transcriptomics. SPR does not require additional labelling of oligonucleotides, and assuming the system is optimised, affinity and kinetic data can be collected simultaneously. In addition, because SPR can be used to screen for DNA oligonucleotides in solution, large numbers of sequences can be screened faster if oligonucleotides are screened in groups rather than by individual sequences.

7.2.7 Mutagenesis and study of chimeric ArpA regulators

Primers for the generation of mutations in AvaL1 LBP residues (F89H, Y90A, Y90F, Y150A, Y150F and Q136E) were designed and ordered (sequences available in the Electronic Supplementary Materials), but mutagenesis was not performed. This could be performed in the future to better understand AHFCA binding in the AvaL1 LBP. Residues in the MmfR DBD which are predicted to form interactions with DNA are also potential mutagenesis targets of future interest. For example, the R48 residue in MmfR may influence DNA-binding if it is introduced into other MmfR homologues. Other mutagenesis studies which could be done include the manipulation of the dimer interface; primers for an $AvaL1_{S185C}$ mutant were also ordered, with an attempt to replicate the cysteine bridge found in the crystal structure of CprB.

Taking a view towards synthetic biology, one could see the individual helices of ArpA-subfamily repressors as modular components, which could be swapped with one another to alter the behaviour of the chimeric protein, such as selecting for different ARE targets using the DBD from one repressor and selecting for a different effector using the LBD from another repressor. This would not be limited to just ArpA-subfamily repressors; one could in theory use components from any member of the TetR-family of repressors. Unusual secondary structures not found in ArpA-subfamily repressors may also be of interest in discovering new uses of the repressor; for example, what would happen if the 10^{th} helix/helix-loop regions found in CprB and TylP were introduced into MmfR? These all lead to a high number of possible permutations of synthetic repressors for development of novel regulatory systems for applications in biotechnology.

8 Appendices

8.1 Appendix A

Table 8-1 - Genes encoded within the methylenomycin (*mmy*) gene cluster of *S. coelicolor* A3(2).

Locus tag	Protein	Length (aa)	Proposed function
SCP1.228c	MmyF	174	Oxidoreductase
SCP1.229c	MmyY	153	Hypothetical protein
SCP1.230	MmyB	313	DNA-binding protein
SCP1.231	MmyQ	217	NADP-dependent oxidoreductase
SCP1.232	MmyE	387	Enoyl reductase
SCP1.233	MmyD	339	AvrD-like protein
SCP1.233a	MmyX	234	ATG/GTP-binding protein
SCP1.233b	MmyC (fabH5)	332	3-oxoacyl-[acyl-carrier-protein] synthase 3 protein 5
SCP1.234	MmyA	82	ACP
SCP1.235	MmyP	232	HAD-like phosphatase
SCP1.236	MmyK	206	ATP/GTP-binding protein
SCP1.237c	MmR	475	Methylenomycin A resistance protein
SCP1.238	MmyJ	111	ArsR-family transcriptional regulator
SCP1.239c	MmyG	393	Oxidoreductase
SCP1.240c	MmyO	373	Monooxygenase
SCP1.241c	MmyT	269	Thioesterase
SCP1.242c	MmfR	214	TetR-family transcriptional regulator
SCP1.243	MmfL	353	AfsA homologue
SCP1.244	MmfH	400	Oxidoreductase
SCP1.245	MmfP	265	Phosphatase
SCP1.246	MmyR	203	TetR-family transcriptional regulator

Table 8-2 - Genes encoded within the gaburedin (*gbn*) gene cluster of *S. venezuelae*.

Locus tag	Protein	Length (aa)	Proposed function
SVEN_4179	GbnA	558	Glutamate decarboxylase
SVEN_4180	GbnB	532	Urea synthetase
SVEN_4181	GbnC	354	Gaburedin exporter
SVEN_4182	SgnR	200	TetR-family transcriptional regulator; MmfR homologue
SVEN_4183	SgnL	345	AfsA homologue; MmfL homologue
SVEN_4184	SVEN_4184	98	LysR-family transcriptional regulator
SVEN_4185	SgnP	257	Phosphatase; MmfP homologue
SVEN_4186	SVEN_4186	75	Hypothetical protein
SVEN_4187	GbnR	206	TetR-family transcriptional regulator; MmyR homologue
SVEN_4188	SgnH	405	Hydrolase; MmfH homologue

Table 8-3 - Genes encoded within the scleric acid (*scl*) gene cluster of *S. sclerotialus*.

Locus tag	Protein	Length (aa)	Proposed function
IG96_RS0135470	SclP	213	PPTase
IG96_RS0135475	SclQ4	386	QncL – Acyltransferase catalytic domain
IG96_RS0135480	SclQ3	338	QncL – Transketolase C-terminal domain
IG96_RS0135485	SclQ2	305	ThDP binding domain
IG96_RS0135490	SclQ1	76	Acyl carrier protein
IG96_RS0135495	SclM1	210	TetR-family transcriptional regulator; MmfR homologue
IG96_RS0135500	SclM2	338	AfsA homologue; MmfL homologue
IG96_RS0135505	SclL	90	LysR-family transcriptional regulator
IG96_RS0135510	SclM3	228	Phosphatase; MmfP homologue
IG96_RS0135515	SclM4	196	TetR-family transcriptional regulator; MmyR homologue
IG96_RS0135520	SclM5	378	Oxidoreductase; MmfH homologue
IG96_RS37880	SclN	1083	NRPS [C-A-PCP]
IG96_RS0135530	SclT	243	Thioesterase
IG96_RS0135535	SclA	633	Anthranilate synthase
IG96_RS0135540	SclD	405	DAHP synthase
IG96_RS0135545	SclI	220	Isochromatase
IG96_RS0135550	SclG	439	ATP-grasp family enzyme
IG96_RS0135555	SclE	424	MFS transporter

Table 8-4 - Genes encoded within the azoxy compound gene cluster of *S. avermitilis*.

Locus tag	Protein	Length (aa)	Proposed function
Saverm_2263		196	TetR-family transcriptional regulator
Saverm_2264		522	Transmembrane efflux protein
Saverm_2265		67	Hypothetical protein
Saverm_2266	AvaC	229	Phosphatase; MmfP homologue
Saverm_2267	AvaB	407	Hydrolase; MmfH homologue
Saverm_2268	AvaL2	206	TetR-family transcriptional regulator; MmyR homologue
Saverm_2269	AvaA	345	AfsA homologue; MmfL homologue
Saverm_2270	AvaL1	218	TetR-family transcriptional regulator; MmfR homologue
Saverm_2272		519	MmgE/PrpDfamily protein
Saverm_2273		122	Isomerase
Saverm_2274		319	Diacylglycerol kinase
Saverm_2275		489	Transmembrane efflux protein
Saverm_2276	fabH7	339	3-oxoacyl-ACP synthase III
Saverm_2277		255	Thioesterase
Saverm_2278		367	F420-dependent dehydrogenase
Saverm_2279		534	Long-chain-fatty-acid-CoA ligase
Saverm_2280	Pks3-1	633	Modular polyketide synthase
Saverm_2281	Pks3-2	1253	Modular polyketide synthase
Saverm_2282	Pks3-3	112	ACP
Saverm_2283		543	Aldehyde dehydrogenase
Saverm_2284		379	Isobutylamine N-hydroxylase
Saverm_2285	RocD3	427	Ornithine aminotransferase
Saverm_2286		160	Cyclase/dehydrase
Saverm_2287		178	Integral membrane protein
Saverm_2288		184	NADPH-flavin oxidoreductase
Saverm_2289		327	Hypothetical protein
Saverm_2290	FabH	335	3-oxoacyl-ACP synthase 3
Saverm_2291	FabC3	76	ACP
Saverm_2292	FabF	406	3-oxoacyl-ACP synthase II
Saverm_2293		506	Amino acid permease
Saverm_2294	SerS2	427	Seryl-tRNA synthetase 2
Saverm_2295	FabI2	260	Enoyl-ACP reductase [NADH]
Saverm_2296	FdxC	108	Ferredoxin
Saverm_2297		469	MFS transporter
Saverm_2298	FadS2	365	Fatty acid desaturase
Saverm_2299		66	Hypothetical protein
Saverm_2300		331	Integral membrane lysyl-tRNA synthetase
Saverm_2301		265	SARP-family transcriptional regulator
Saverm_2302		143	Membrane protein
Saverm_2303	MmdA1	514	Methylmalonyl-CoA carboxyltransferase
Saverm_2304		61	Hypothetical protein
Saverm_2305		132	Enamine deaminase
Saverm_2306	FadE10	375	Acyl-CoA dehydrogenase
Saverm_2307		594	Acyl-CoA synthetase
Saverm_2308		280	PaaX-family transcriptional regulator

8.2 Appendix B

sclM1co

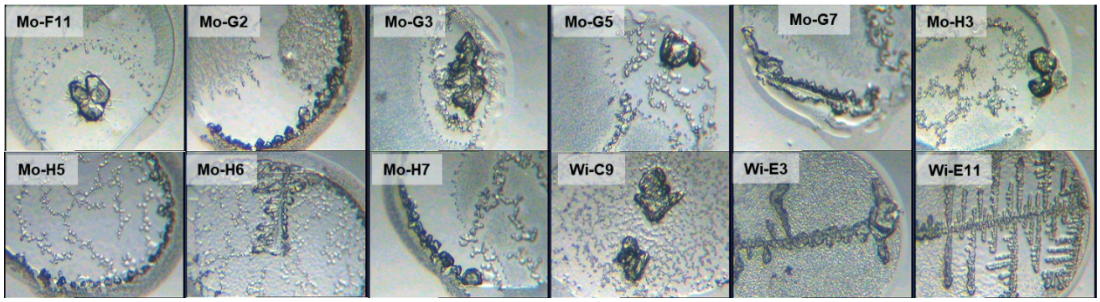
CACCATGGATAGCAAAGCAGCAACCCATGGTCCGGTTGCACATAGCCGTCGTGAACCGAAACAAGAACGTGCCCGTCAGACCAAACGTCTGGTTCTGCG
TACCGCAGCCGAAGCATTTAAAGAACATGGTTATCCGGCAGTTACCCTGCAGGCAGTTGCAGAACGTGCAAGCATGACCAAAGGCGCAGTTTATTTTCA
TTATGCCAATAAAGAAGCACTGGCCGTTGCAGTTATTATGGAACATTATGCACGTTGGGAACCGCTGGTTAGCGAAGTTCGTAGTCGCGGTCTGCCTCC
GCTGGAAACCCTGCTGGCCGTTCTGGAAGGCGCAGCGGAAGCCTTTCGTGATGATGTTATGGTTCAGGCAGGCGCACGTCTGCAGATTGAACGTAGCCT
GATTAAAGCCGATCTGCCGGTTCCGTATGTTGGTTGGCAAGAACTGCTGACCAGTCTGATTACCGAAGCACGTGATGCAGGTCAGCTGCGTGCAGATGT
TGATCCGGCAGCACTGGCACGTGTTGTTGTTGCAGCATTTTTTGGTAGCCAGCATATTAGTGATGTTCTGAATGGTCGTGCAGATCTGATGCAGCGTCA
TGGCGAACTGATTGAAGCAGTTTTTCGCGGAGCAATGCCGTAA

sclM4co

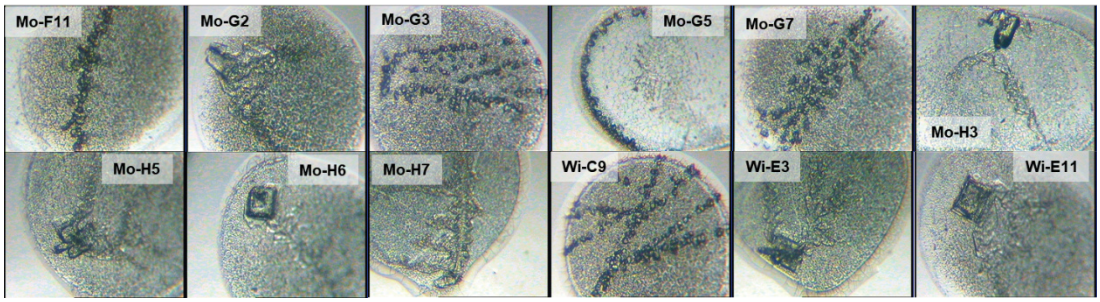
CACCATGGTTAAACAGGATCGTGCACGTCGTACCCATGCACTGGTTCTGGATGCAGCAGCAGCCGAATTTAGCGCACTGGGTTTTGCAAGTACCAATCT
GCAGGTTGTTGCAGCACGTACCGGTCTGACCAAAGGCGCACTGTATGGTCATTTTCCGAGCAAAAGTGCACTGGCAGTTGAAATTACCCGTCAGTTTGA
AGAAAGCTGGCGTGAACTGCTGCGTTGTGTTGATAAAGAAACCGTTCTGCCGCTGACCGCACTGCATGCACTGCTGCTGGGTCTGGCACAGAAAATTCA
GACCGATGTTCGTTTTACCGCAGGTCTGCGTCTGGTTAGCGAAGAGGCACGTGCAAAAGGCGAAGTTCCGGAAGTTCTGACCGAACTGCGCGGAGCAAT
GCTGCATCTGGCCCAGCAGGCCCAGAAAGATGGCGATATTGATGCAGGTATTGGCGCAGAACCGCTGAGCCATCTGCTGCTGAGCCTGGTACTGGGTAT
TCATCATACAACCAGCGCAGGCGATAGCGATAGCGTTTATCAGCGTGTTGTTAGCGCATGGGAGCTGCTGCTGCCTCTGATGCGTCGTCCGGTTCGTTA
A

Figure 8-1 – GeneArt® String™ DNA fragments used for cloning of *E. coli* codon-optimised *S. sclerotialus* *sclM1* and *sclM4* into *E. coli* TOP10 using the pET151/D-TOPO® cloning system. DNA was suspended in 10 mM Tris pH 8.5 buffer and delivered dry. Forward and reverse primers for each sequence are shown. Additional parameters prevent the inclusion of CACC, CCAC, GTGG and GGTG within the sequences to stop misorientation within the plasmid.

A



B



C

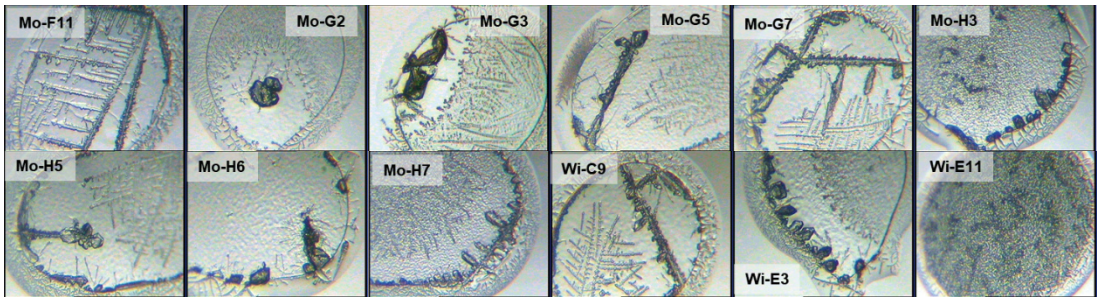


Figure 8-2 – Initial hits from crystallisation trials for unbound AvaL1, AvaL2 and SgnR after 31 days. (**A**) AvaL1 at 26 mg mL^{-1}, (**B**) AvaL2 at 22 mg mL^{-1} and (**C**) SgnR at 31 mg mL^{-1}. Droplets have crystallant:protein volumes of 100:100 nL. Screens were set up in a HT-96 well plate and screened on a Mosquito LCP. Mo = Morpheus and Wi = Wizard Classic I and II. Images were collected on 14th March 2016.

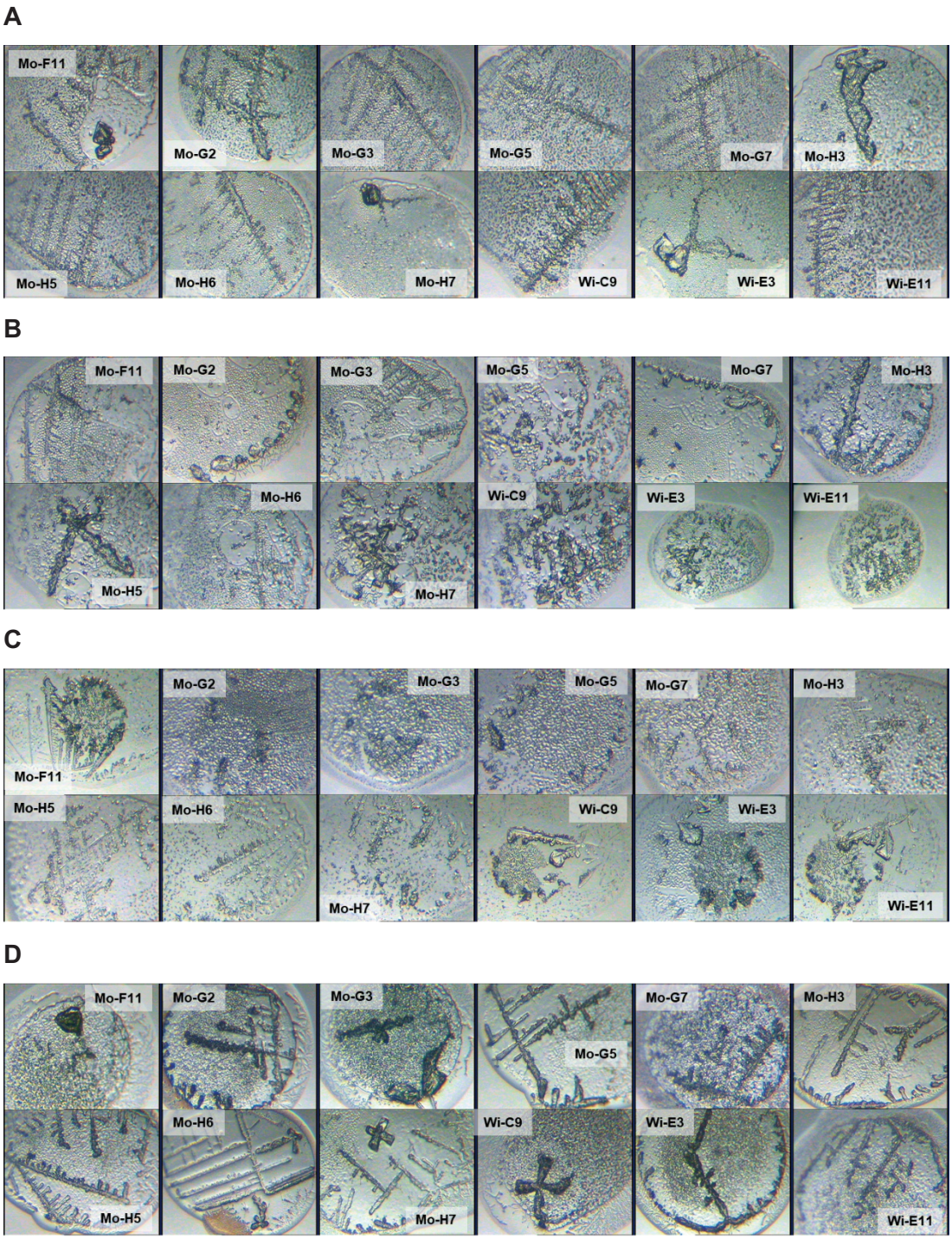


Figure 8-3 – Initial hits from crystallisations trials of AvaL1 and SgnR co-screened with ligands after 31 days. 26 mg mL^{-1} AvaL1 was co-screened with (**A**) AHFCA1 and (**B**) AHFCA5, while 31 mg mL^{-1} SgnR was co-screened with (**C**) SCB1 and (**D**) AHFCA5. All droplets have molar ratios of 1:2 protein:ligand, at crystallant:protein/ligand volumes of 100:100 nL. Screens are set up in a HT-96 well plate, with screens set up on a TTPLabtech Mosquito LCP. Mo – Morpheus and Wi = Wizard Classic I and II. Images were collected on 14th March 2016.

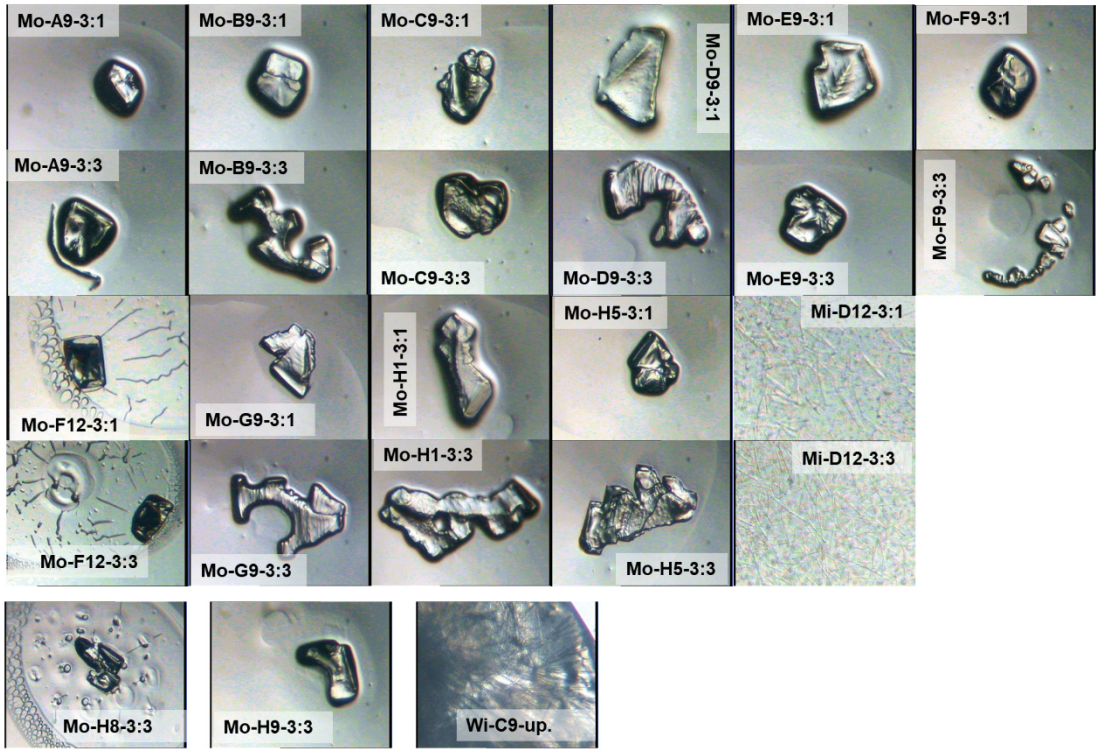


Figure 8-4 – Hits from crystallisation trials for 31 mg mL^{-1} AvaL1 at crystallant:protein drop ratios of 300:100 and 300:300 nL after 26 days. Screens are set up in a HT-96 well plate, with screens set up on a TTPLabtech Mosquito LCP. Mo = Morpheus, Wi = Wizard Classic I and II, and Mi = MIDAS. Images were collected on 16th August 2016.

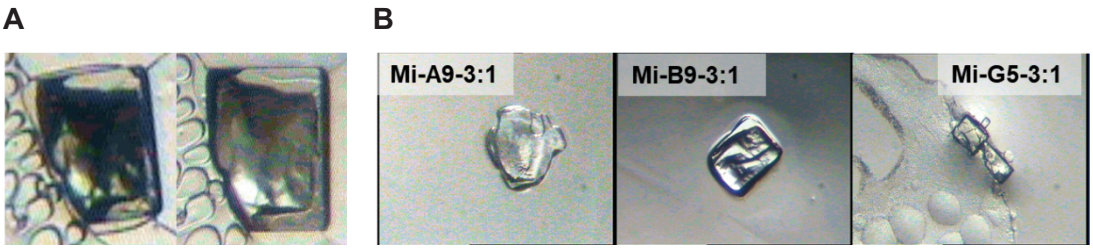


Figure 8-5 – Additional precipitation events for screens performed as shown in Figure 8-4. (**A**) Crystal screening hit from AvaL1 (16 mg mL^{-1}) for the Morpheus condition F12 (0.12 M Monosaccharides, 0.1 M Buffer System 3, pH 8.5, 50 % v/v Precipitant Mix 4) on September 12th, 2016 (left) and 2 weeks later on September 26th (right) at droplet ratio of 300:100 nL crystallant:protein. (**A**) Crystal screen results for AvaL1 from the MIDAS™ kit collected on September 26th, 2016. Screens are set up in a HT-96 well plate, with screens set up on a TTPLabtech Mosquito LCP. Drop ratios are crystallant:protein of 100 nL.

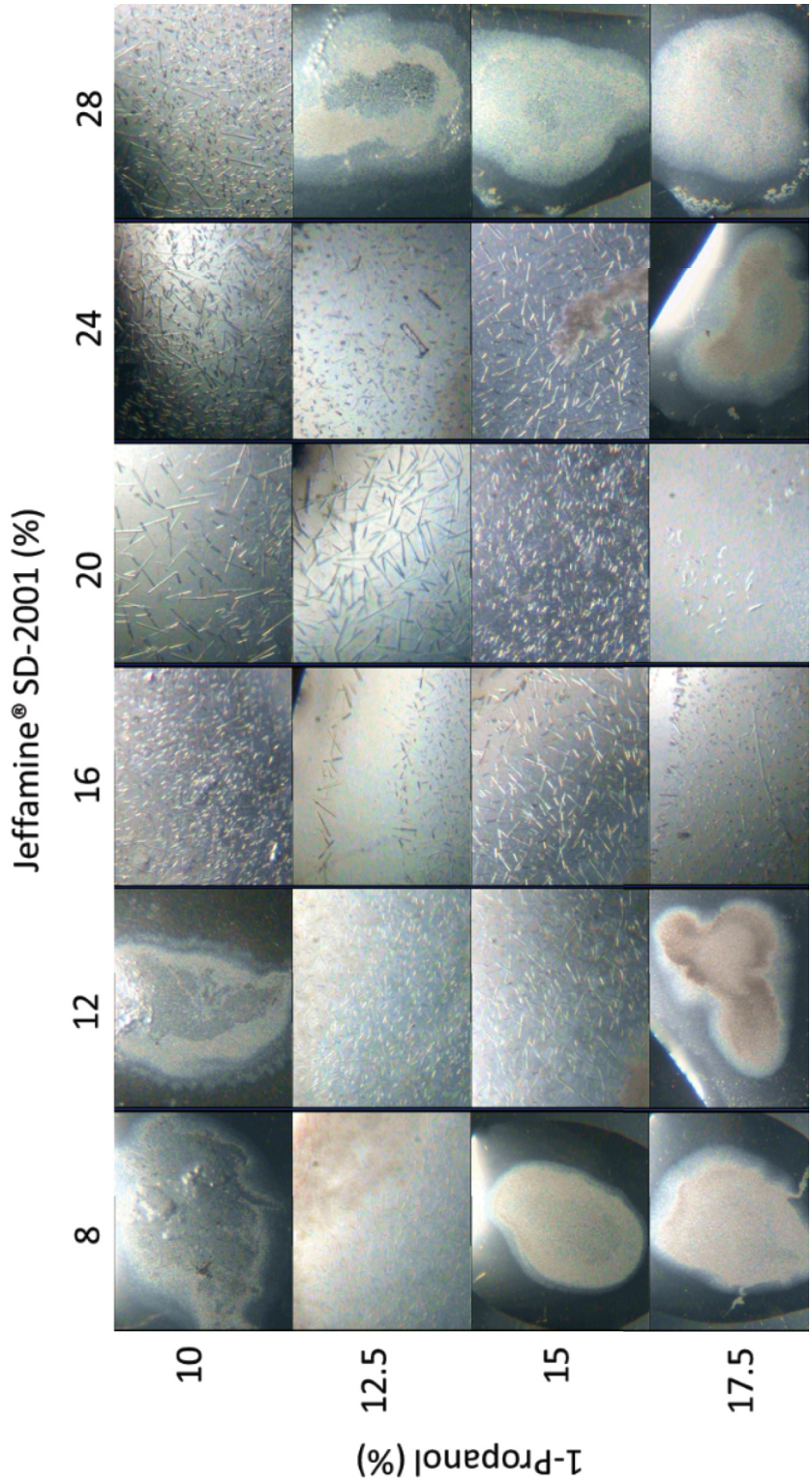


Figure 8-6 – Upscaled optimisation grid of 1:1 µL crystallant:protein crystallisation trials for 15.8 mg mL^{-1} AvaL1 in MIDAS™ D12 crystallant conditions (stock condition 20 % v/v Jeffamine® SD-2001, 15 % v/v 1-Propanol), 8 days after initialisation (15th September 2016) varying Jeffamine® SD-2001 from 8-28 % v/v, and 1-Propanol from 10 to 17.5 % v/v.

8.3 Appendix C

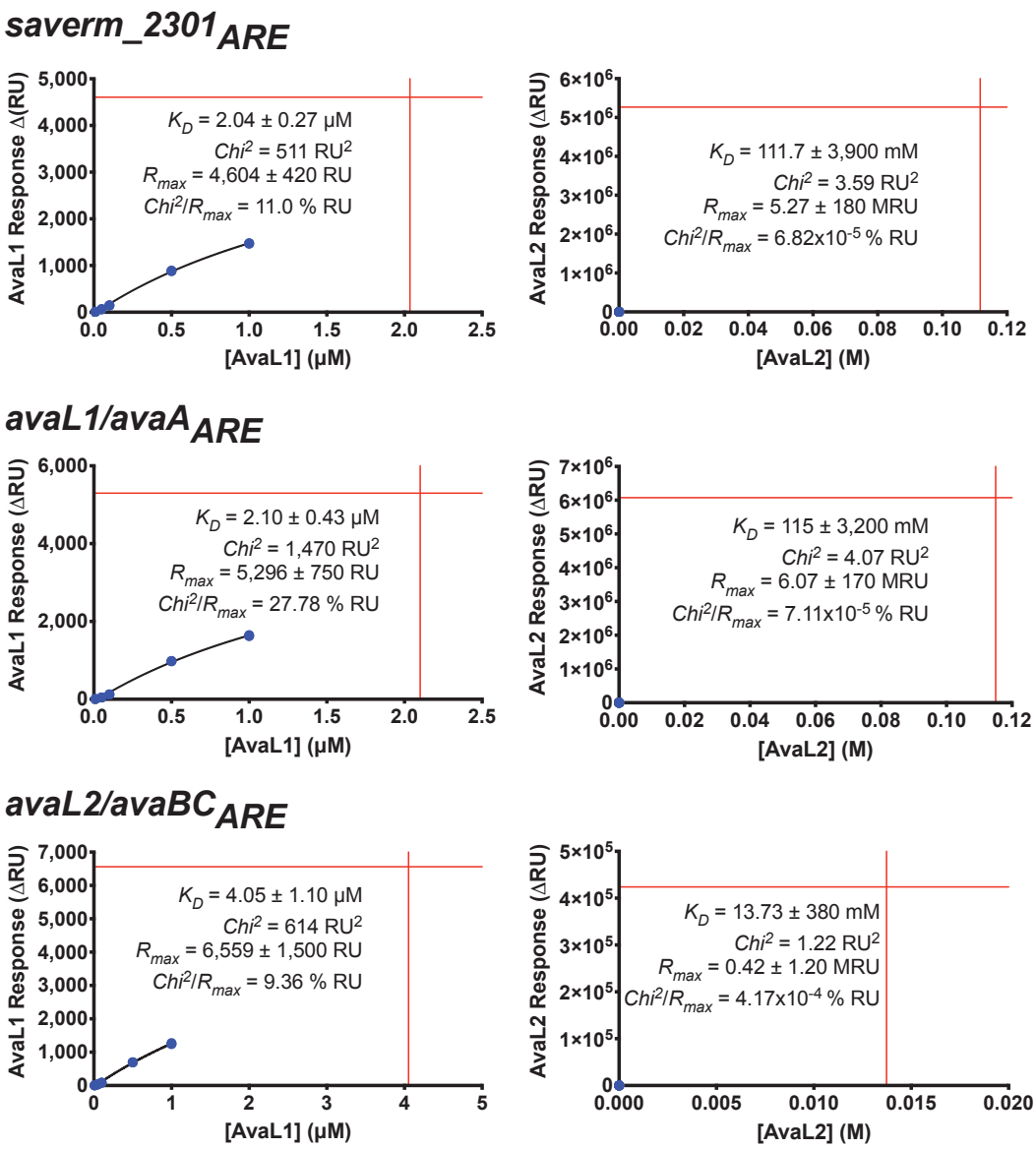


Figure 8-7 – Fitted steady state saturation curves for the pilot binding of AvaL1 and AvaL2 to SA-immobilised *S. avermitilis* AREs. AvaL1/AvaL2 at concentrations of 10 nM, 50 nM, 100 nM, 500 nM and 1000 nM were injected over flows cells functionalised with (from top to bottom) *saverm_2301*$_{ARE}$, *avaL1/avaA*$_{ARE}$ and *avaL2/avaBC*$_{ARE}$. AvaL1 binding is on the left, and AvaL2 on the right. The red vertical line indicates the determined K_D value, and the red horizontal line shows the fitted R_{max} value.

Table 8-5 – Unconstrained Kinetic Model fitting values for the pilot binding of AvaL1 to SA-immobilised *S. avermitilis* AREs. Units for R_{max}, Chi^2 and Chi^2/R_{max} are in RU, RU^2 and % respectively. Kinetic models are as follows: L = 1:1 Langmuir, BA = bivalent analyte, HL = heterogenous ligand, 2S = two state. Asterisks indicate models with which refractive index (RI) offsets are constrained to 0.

	***saverm_2301*$_{ARE}$**			***avaL1/avaA*$_{ARE}$**			***avaL2/avaBC*$_{ARE}$**		
Model	**Chi²**	R_{max}	$\frac{Chi^2}{R_{max}}$	**Chi²**	R_{max}	$\frac{Chi^2}{R_{max}}$	**Chi²**	R_{max}	$\frac{Chi^2}{R_{max}}$
L	3,870	994 ± 30	389.2	3,800	2,320 ± 47	163.8	2,520	1,442 ± 45	174.8
L*	8,090	4,053 ± 11	199.6	6,150	4,568 ± 9	134.6	3,650	9,988 ± 72	36.5
BA	3,020	1,632 ± 31	185.0	895	5,141 ± 9	17.4	139	41,760 ± 3,400	0.3
BA*	562	7,739 ± 89	7.3	1,110	9,719 ± 170	11.4	583	11,100 ± 12	5.3
HL	186	990 ± 44	1.9	1,150	743 ± 3	33.7	136	58,170 ± 1,800	0.2
		8,786 ± 210			2,671 ± 3			948 ± 80	
HL*	507	989 ± 71	8.9	1,110	5,775 ± 57	15.6	516	1,219 ± 81	2.6
		4,723 ± 51			1,324 ± 68			18,410 ± 86	
2S	**147**	**9,095 ± 28**	**1.6**	**229**	**6,760 ± 82**	**3.4**	**137**	**43,210 ± 2,700**	**0.3**
2S*	492	5,708 ± 57	8.6	1,080	7,500 ± 120	14.4	502	18,100 ± 93	2.9

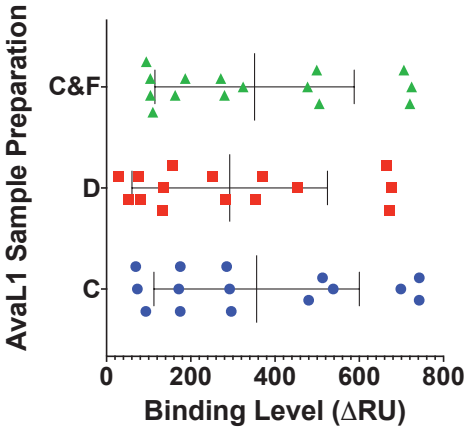
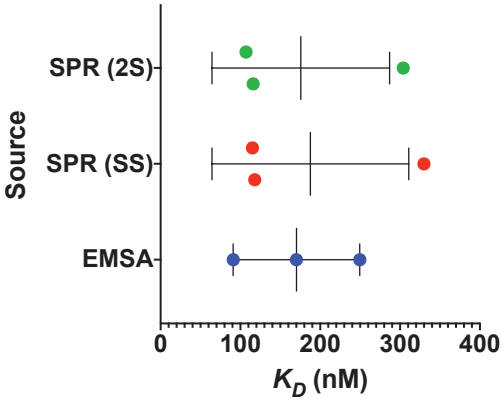
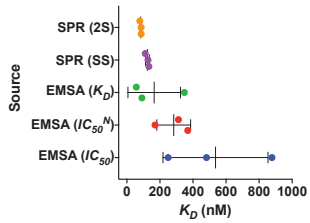


Figure 8-8 – Box and Whisker plot showing the similarity in *BL* values of AvaL1 binding to *avaL1/avaA*$_{ARE}$ using differing sample preparation methods across. C = centrifugation, D = dialysis, and C&F = centrifugation and freezing. Shown are mean *BL* values and with standard deviation. A Levene's test ($W(2, 42) = 0.11$, $p = 0.90$) indicates homoscedasticity (equal variance between datasets). ANOVA between the three sources indicate the population means are equal ($F(2, 42) = 0.34$, $p = 0.72$).

Figure 8-9 – Box and Whisker plot showing the differences in mean affinity of AvaL1 binding to *avaL1/avaA*$_{ARE}$ as derived from gel shift assays (IC_{50}, IC_{50}^N and K_D) and SPR experiments optimising flow cell surface density (steady state, SS, and two state, 2S). A Levene's test indicates that the variance is homoscedastic ($W(2, 6) = 0.85$, $p = 0.47$), and ANOVA indicates the mean values are equal ($F(2, 6) = 0.02$, $p = 0.98$).



A



B

	EMSA (nM)			**SPR K_D (nM)**	
ARE	IC_{50}	IC_{50}^N	K_D	**SS**	**2S**
saverm_2301$_{ARE}$	482	312	348	133	78
avaL1/avaA$_{ARE}$	250	170	91	111	86
avaL2/avaBC$_{ARE}$	878	368	57	128	87
Mean (with SD)	537 ± 383	283 ± 102	199 ± 146	124 ± 12	84 ± 5
t(2)	2.30	3.01	0.46	-	4.64
p	**0.15**	**0.10**	**0.68**	-	**0.04**

Figure 8-10 – Hypothesis testing for the difference in the mean affinity values for AvaL1 binding to all three endogenous *S. avermitilis* AREs derived from EMSA and SPR experiments. (**A**) Box and Whisker plot for each source with mean values and standard deviation shown. (**B**) Values plotted in **A** with post hoc paired Student's *t* tests between each dataset and the SS dataset. SS = steady state and 2S = two state models.

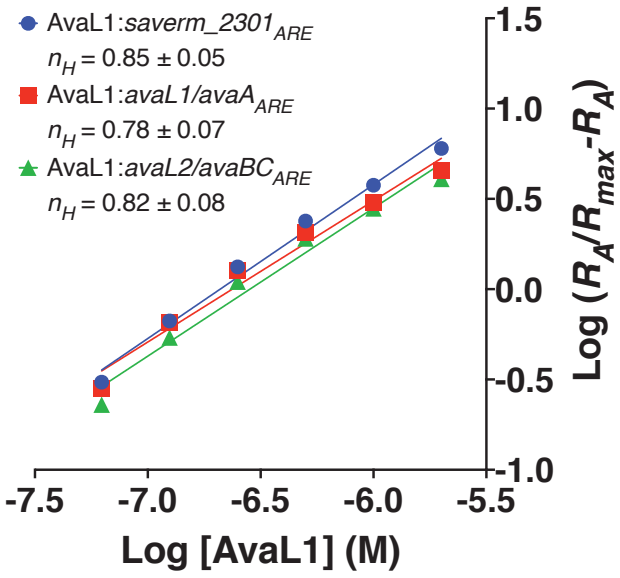
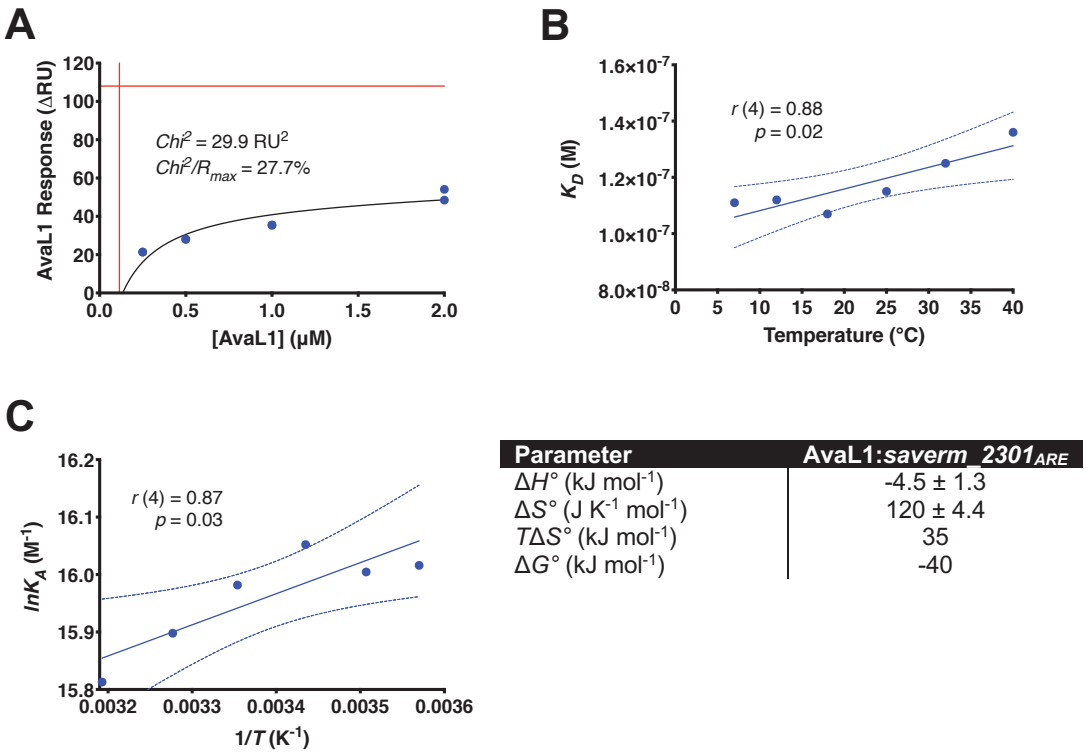


Figure 8-11 – Linear Hill plots for AvaL1 binding to SA-immobilised *S. avermitilis* AREs. R_A values and R_{max} are taken from the data in Figure 4-7. For all linear regressions $r(4) = 0.99$, $p < 0.01$. n_H = Hill coefficient.



Parameter	AvaL1:*saverm_2301*$_{ARE}$
$\Delta H°$ (kJ mol^{-1})	-4.5 ± 1.3
$\Delta S°$ (J K^{-1} mol^{-1})	120 ± 4.4
$T\Delta S°$ (kJ mol^{-1})	35
$\Delta G°$ (kJ mol^{-1})	-40

Figure 8-12 –Thermodynamic analysis of AvaL1 binding to *saverm_2301*$_{ARE}$ with R_{max} constrained to TR_{max} of 108 RU. (**A**) Steady state saturating curve fitted to 0.25 μM to 2 μM (two-fold serial dilution) AvaL1 binding to *saverm_2301*$_{ARE}$ at 25 °C, with R_{max} and K_D indicated in red horizontal and vertical lines respectively. (**B**) K_D plotted against temperature (over a range of 7 °C to 40 °C). (**C**) van't Hoff plot with table of thermodynamic parameters. $T\Delta S°$ and $\Delta G°$ are calculated assuming T = 298.15 K (25 °C). For **B** and **C** 95 % confidence intervals were also plotted.

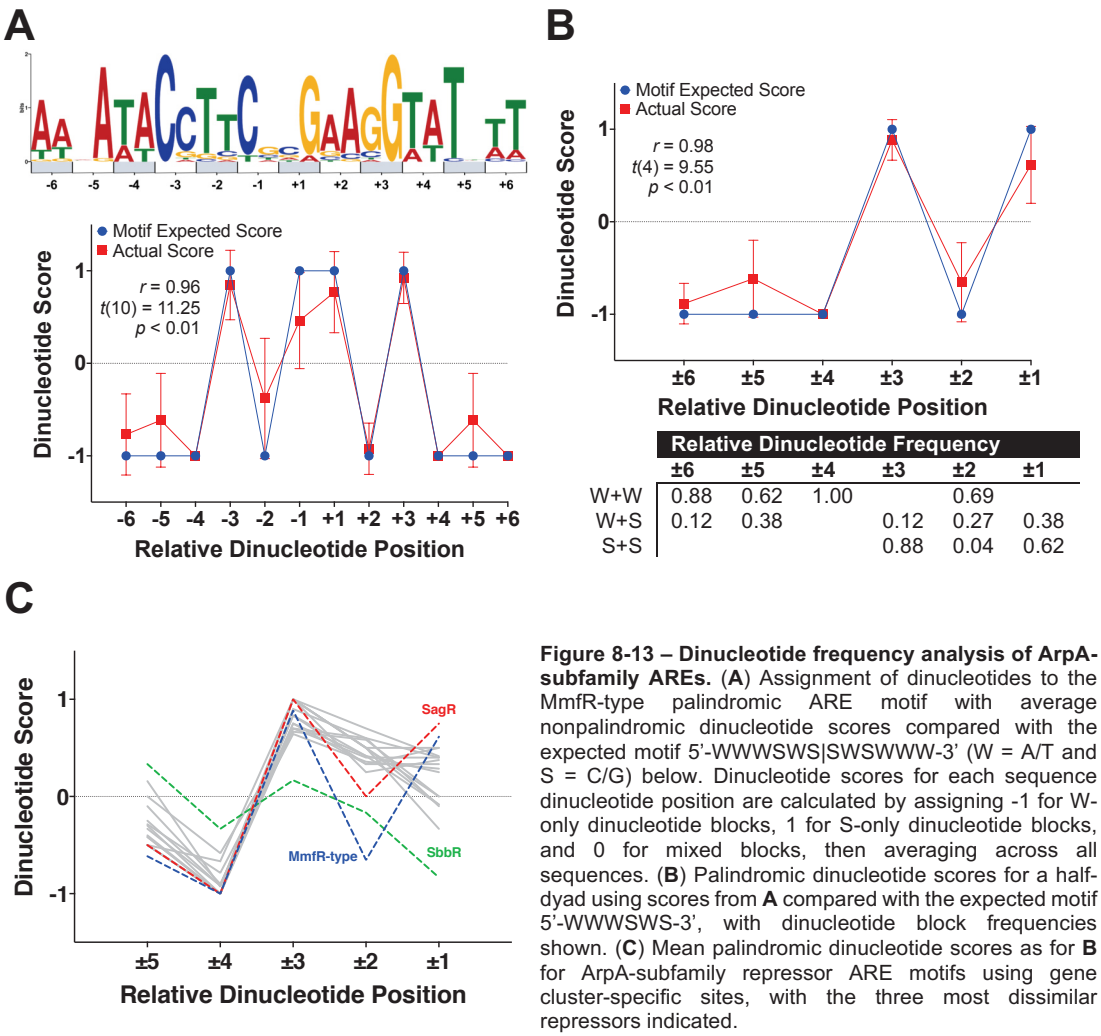
Table 8-6 – (next page) Tag/Fusion composition of TFRs for which affinity of binding to DNA has been determined. N-terminal tag amino acid sequences are given for each protein as assayed, post-cleavage in the case of tags being removed at the indicated cleavage site (protease recognition site is underlined, with cleavage site indicated by \). The number of amino acids (aa) in the extension are also indicated. References can be found in the text of Chapter 4.4.2.

TFR	Expression Vector	Tag	N/C- tag Terminal	Tag cleaved?	Protease cleavage site type	N-Terminal tag sequence (post-cleavage)
AvaL1	pET151	His_6	N	No	TEV	MH_6GKPIPNPLLGLDSTENLYFN\GIDPFT (33 aa)
AcrR	pET15b	His_6	C	No		-
EthR	pET15b	His_6	N	No	Thrombin	MGSSH$_6$SSGLVPR\GSH (20 aa)
MtrR	pET15b	His_6	N	No	Thrombin	MGSSH$_6$SSGLVPR\GSH (20 aa)
RamR	pET15b	His_6	N	No	Thrombin	MGSSH$_6$SSGLVPR\GSH (20 aa)
AtrA	pET16b	His_{10}	N	No	Factor Xa	MGH$_{10}$SSGHIE\GRH (21 aa)
CifR	pET16b	His_{10}	C	Yes	HRV C3	-
FadR	pET22b	His_6	C	No	-	-
ArpA	pET26b	His_6	C	No	-	-
CprB	pET26b	His_6	C	No	-	-
ActR	pET28a	His_6	N	No	Thrombin	MGSSH$_6$SSGLVPR\GSH (20 aa)
AibR	pET28a	His_6	N	Yes	Thrombin	\GSHMAS (6 aa)
AlkX	pET28a	His_6	N	No	Thrombin	MGSSH$_6$SSGLVPR\GSHMASMTGGGGMGRGSE (35 aa)
Fad35R	pET28a	His_6	N	Yes	Thrombin	\GSHMAS (6 aa)
SCO7222	pET28a	His_6	N	No	Thrombin	MGSSH$_6$SSGLVPR\GSH (20 aa)
XdhR	pET28b	His_6	N	No	Thrombin	MGSSH$_6$SSGLVPR\GSH (20 aa)
TtgR	pET29a	-	-	-	-	-
KstR	pET30a	His_6	N	Yes	Thrombin	\GSHM (4 aa)
TetR/revTetR	pET3c	-	-	-	-	-
AcnR	pET-TEV	His_6	N	Yes	TEV	\GH (2 aa)
SmcR	pHis-Parallel1	His_6	N	No	TEV	MSYYH$_6$DYDIPTTENLYFE\GAM (26 aa)
C. glutamicum AmtR	pMAL-C2	MBP	N	No	Factor Xa	MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSNNNNNNNNNNLGIE\GRISEFGS (393 aa)
CmeR	pQE-30	His_6	N	No	-	MH_6 (7 aa)
QacR	pSK5210	His_6	C	No	-	-
M. smegmatis AmtR	pYUB28b	His_6	C	No	-	-

Table 8-7 – Hill coefficients (n_H) reported for TFRs binding DNA as pairs of dimers.

TFR:*DNA* complex	Source	n_H	Ref.	Notes
ActR:*actR/actA*$_{ir}$	EMSA	~1	(222)	
TtgR:*ttgR/ttgA*$_{(wt)}$	ITC	1.64 ± 0.13	(215)	
TtgR:*ttgR/ttgA*$_{(op5)}$	ITC	1.81 ± 0.05	(215)	
AlkX:P_{alkW1}	ITC	1.63 ± 0.03	(216)	
SCO7222:*SCO7222/SCO7223*$_{ir}$	EMSA	1.6	(222)	Dimer of dimers requires confirmation
AcrR:*bplR*	FP	2.06 ± 0.12	(219)	
CprB:*OPB*	ITC, FA/FP	2.08, 1.6 - 2.0	(69)	From ITC $n_H = 2/1 + (K_A/K_A')^{1/2}$
EthR:*ethA-R*$_{ir}$	SPR	1.73	(206)	n_H = 3.46 for EthR binding to two operators

8.4 Appendix D



Relative Dinucleotide Frequency						
	±6	±5	±4	±3	±2	±1
W+W	0.88	0.62	1.00		0.69	
W+S	0.12	0.38		0.12	0.27	0.38
S+S				0.88	0.04	0.62

Figure 8-13 – Dinucleotide frequency analysis of ArpA-subfamily AREs. (**A**) Assignment of dinucleotides to the MmfR-type palindromic ARE motif with average nonpalindromic dinucleotide scores compared with the expected motif 5'-WWWSWS|SWSWWW-3' (W = A/T and S = C/G) below. Dinucleotide scores for each sequence dinucleotide position are calculated by assigning -1 for W-only dinucleotide blocks, 1 for S-only dinucleotide blocks, and 0 for mixed blocks, then averaging across all sequences. (**B**) Palindromic dinucleotide scores for a half-dyad using scores from **A** compared with the expected motif 5'-WWWSWS-3', with dinucleotide block frequencies shown. (**C**) Mean palindromic dinucleotide scores as for **B** for ArpA-subfamily repressor ARE motifs using gene cluster-specific sites, with the three most dissimilar repressors indicated.

Table 8-8 – Additional bioinformatic analyses for ArpA-subfamily ARE motifs. (**A**) % GC content analysis of the 22-bp motifs generated for each ArpA-subfamily repressor**.** *p* values are for an unpaired Student's *t* test comparing the mean % GC content of MmfR with other repressor binding sites, with a null hypothesis of identical mean values. (**B**) Palindromic analyses using Tomtom and MEME motif *E* values. Log *E* value ratio for ArpA-subfamily was estimated due to inability to calculate the log value of a number with such a low negative exponential value; instead ratio was measured directly as the ratio of the exponential values themselves. Asterisks for TFRs indicate gene cluster-specific + pleiotropic AREs, while Asterisks for log *E* ratio values indicate that values are inaccurate because the calculated motifs are not significant.

	A				**B**	
TFR	**Mean % GC**	**Min Site % GC**	**Max Site % GC**	***p***	**Tomtom interpalindromic *p* value**	$\frac{\log E_{nonpalindromic}}{\log E_{palindromic}}$
SbbR*	68.2 ± 18.6	32	01	< 0.01	9.81^{-11}	0.29
ArpA	60.1 ± 5.2	50	73	< 0.01	1.46^{-12}	0.83
AvaR2*	58.4 ± 11.6	36	86	< 0.01	2.91^{-13}	0.85
AvaR1*	58.1 ± 12.5	36	86	< 0.01	2.95^{-14}	0.88
ScbR2*	56.6 ± 9.5	36	73	< 0.01	3.67^{-11}	0.79
ScbR*	55.7 ± 12.3	36	73	< 0.01	4.20^{-13}	0.66
CprB	54.7 ± 9.0	41	68	< 0.01	3.05^{-10}	0.50
JadR2	54.5 ± 6.4	50	59	0.01	2.04^{-14}	-0.08*
FarA	52.8 ± 6.3	50	64	< 0.01	1.03^{-14}	0.80
ScbR2	50.1 ± 7.0	41	59	< 0.01	6.81^{-13}	0.46
JadR3	50.0 ± 7.3	41	59	0.01	1.74^{-13}	0.75
SagR	50.0 ± 7.19	45	55	0.04	1.17^{-13}	-0.02*
Aur1R	49.3 ± 12.6	27	64	0.02	2.30^{-13}	0.75
ScbR	49.2 ± 8.1	41	59	0.01	3.47^{-12}	0.60
TylP	48.7 ± 7.1	41	55	0.04	1.78^{-13}	0.88
SpbR	48.5 ± 4.8	41	55	< 0.01	1.31^{-12}	0.68
AvaR1/AvaR2	48.5 ± 10.1	36	64	0.02	5.98^{-12}	0.81
FarR2	48.2 ± 13.3	27	64	0.05	7.66^{-14}	0.67
AlpZ/AlpW	47.0 ± 7.2	41	55	0.07	1.09^{-11}	0.54
BulR1/BulR2	46.7 ± 4.1	45	55	0.02	4.99^{-12}	0.76
AvaR3	43.7 ± 7.1	36	50	0.19	7.86^{-13}	0.58
SbbR	39.6 ± 13.3	32	55	0.05	7.06^{-9}	0.00*
MmfR-type	39.4 ± 7.5	27	50	-	2.32^{-12}	0.92

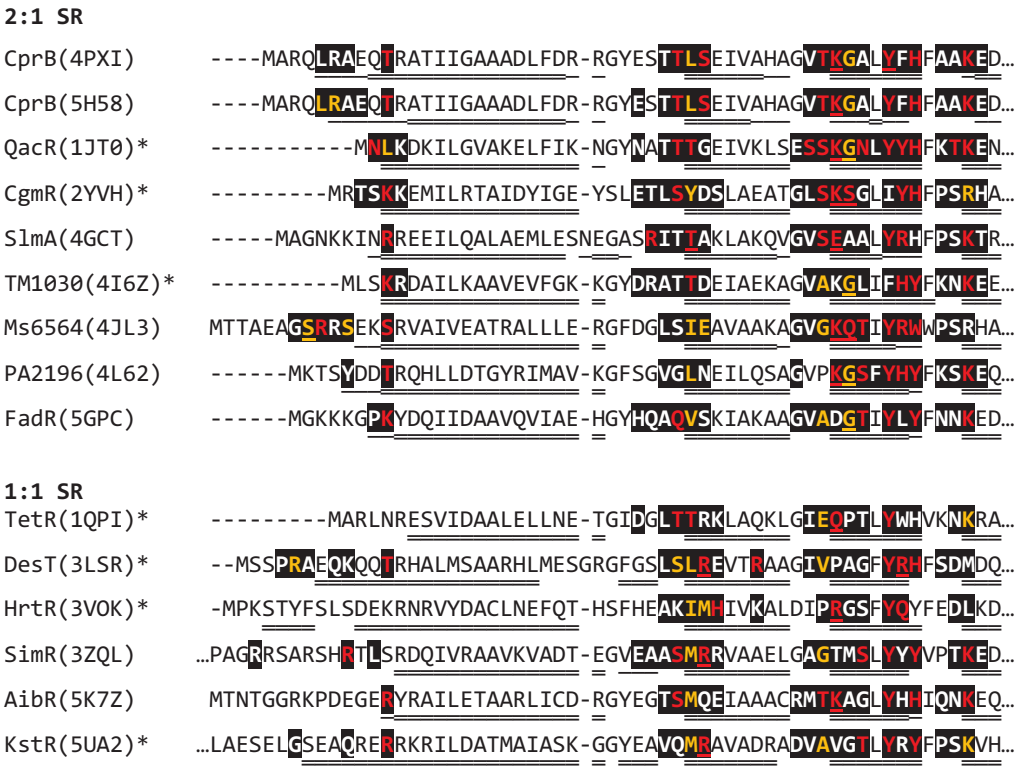


Figure 8-14 – MUSCLE Alignment of the DBD of available TFR/DNA complexes with residues annotated by type. Asterisk indicate TFR/DNA complexes where symmetry is generated from symmetry mates. Amino acids within 5 Å are highlighted in **black**, residues that interact *via* side-chains are **red**, while interactions *via* residue are text-coloured **yellow**. Residues that can interact with nucleotide bases are underlined. Secondary structure annotation is based on consensus secondary structure determination from PDB coordinates using 2Struc[254]. Alpha-helices conserved between all chains is indicated by ═, while ─ indicates non-consensus between chains. PDB identifiers are indicated in parentheses.

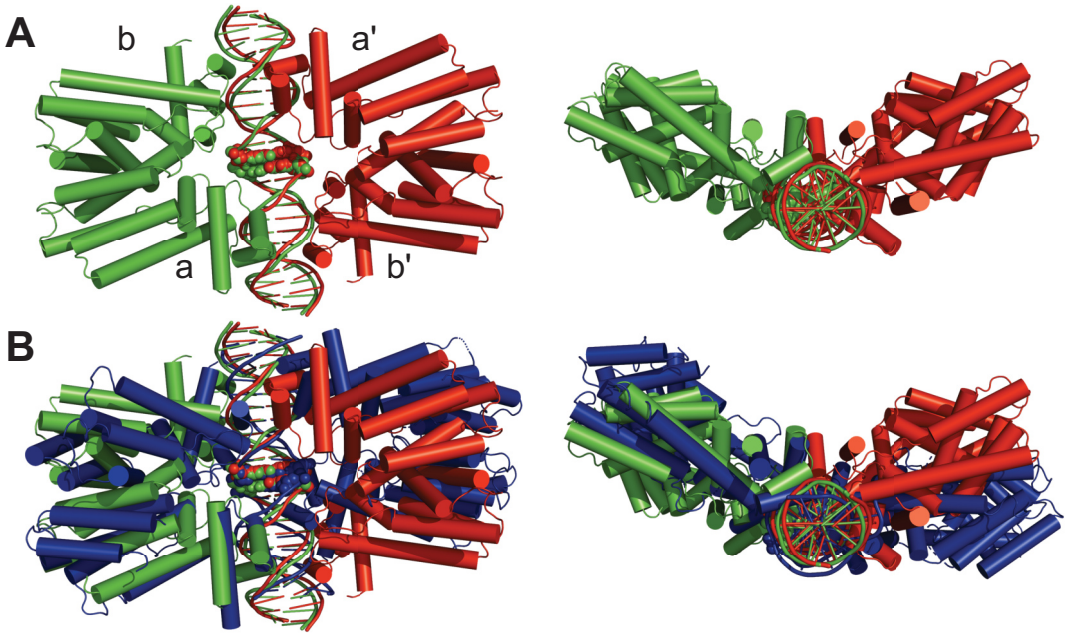


Figure 8-15 – Symmetry model of two MmfR dimers docking to the MmfR-type consensus sequence, based on the docking observed for Figure 5-10. The second MmfR dimer (**red**) was generated from duplication of original single MmfR-docked structure (**green**) and then aligned using Cealign to the complimentary strand position. (**A**) the generation structure viewed from the top and through a -90 turn through the x-axis. (**B**) the same as A, but with the CprB structure (PDB:4PXI) (**blue**) also depicted. Subunits are indicated in lower case letters, with ' indicating symmetry.

8.5 Appendix E

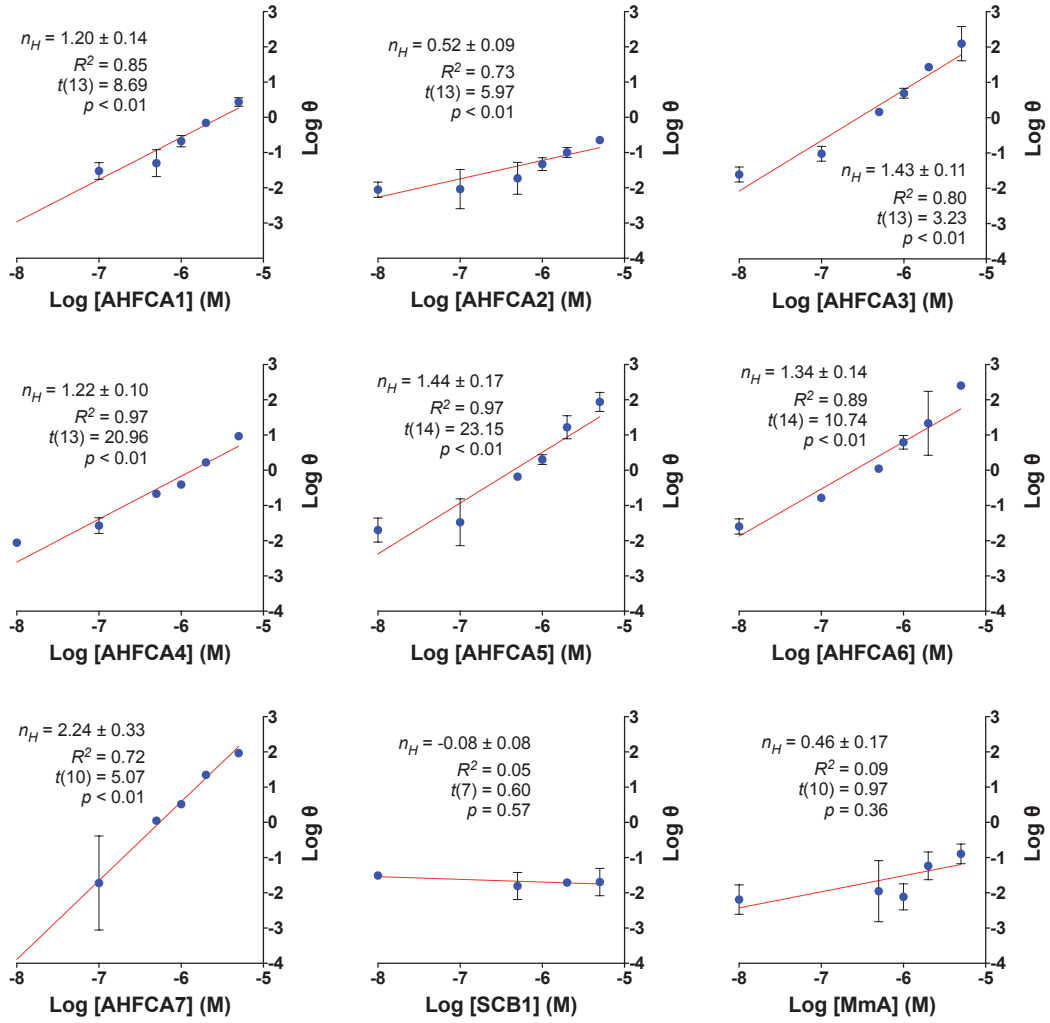


Figure 8-16 – Hill plots for the inhibitory activity of each effector against the DNA-binding of AvaL1. θ is defined as Y/1-Y, where Y is the fraction of AvaL1 not bound to *saverm_2301*$_{ARE}$ in the presence of the effector (Y-inversion of AvaL1 fraction bound to DNA). The Hill coefficient n_H is equal to the slope of the linear regression line. Log θ values for Y values either negative or greater than 1 could not be determined. For all AHFCA molecules n_H = 1.34 ± 0.45 and excluding AHFCA2 n_H = 1.48 ± 0.45.

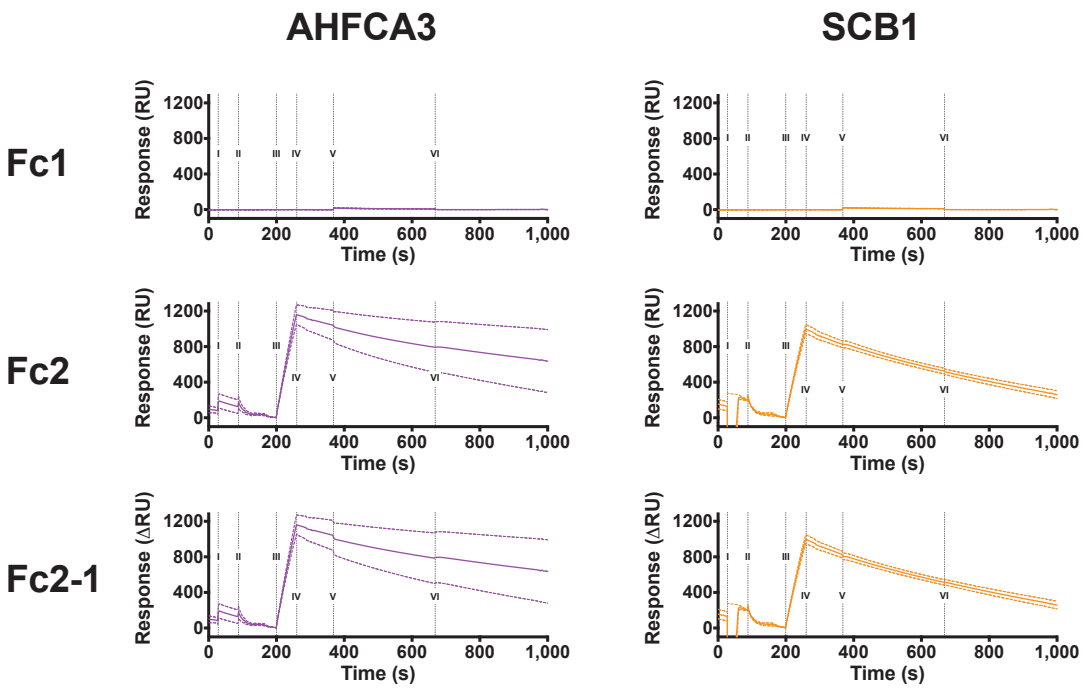


Figure 8-17 – SPR Sensorgrams for the attempted binding of 2 µM AHFCA3/SCB1 to NTA-immobilised AvaL1. Shown are reference Fc1 (no AvaL1), Fc2 with AvaL1 immobilised, and the referenced subtracted Fc2-1. Important injection events are indicated in numerals on the graphs as follows: (**I**) $NiCl_2$ injection, (**II**) running buffer equilibration, (**III**) 200 nM AvaL1 injection, (**IV**) running buffer equilibration, (**V**) effector injection, and (**VI**) running buffer injection. Broken lines indicate SEM of 2 replicates.

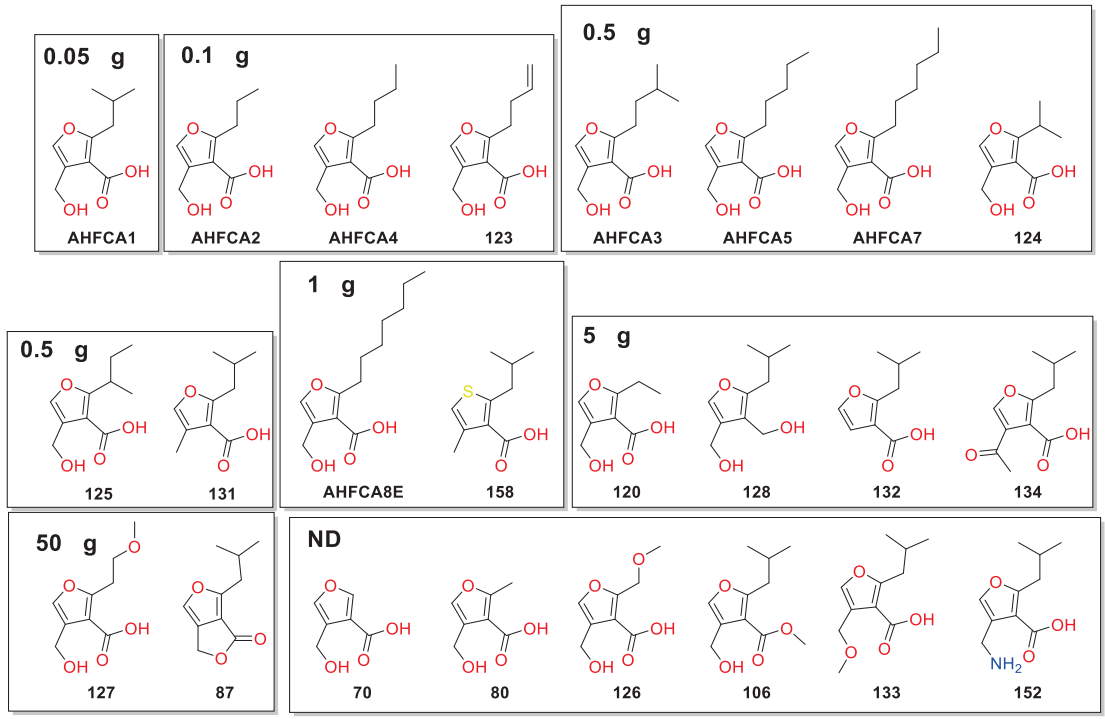


Figure 8-18 – Compounds previously assayed for MmA bioactivity in *S. coelicolor* A3(2) *in vivo* by Dr Malet. Molecules are grouped according to the minimum required amount of inducer to induce a response.

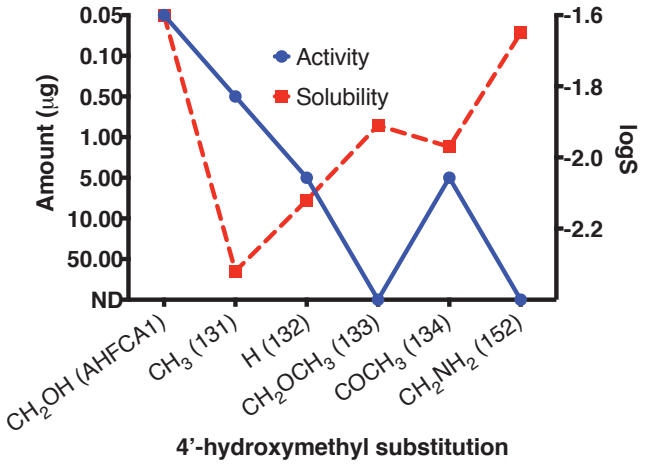


Figure 8-19 – Correlation of *in* vivo methylenomycin bioassays with synthetic 4'-hydroxymethyl substituted AHFCA inducers compared with their theoretical intrinsic solubility (logS). Compound identifiers are indicated in parenthesis (structures in Figure 8-20, Appendix E). There is no significant correlation between the two variables ($r(4) = -0.10$, $p = 0.86$).

Table 8-9 – AutoDock Vina scores for docking of AHFCA analogue compounds shown in Figure 8-18 to the crystal MmfR:AHFCA2 crystal structure. Shown also are the mean docking scores of each class for all molecules minus the MeOH-type molecules. Names in parentheses are compound labels as used by Dr Malet.

Compound	(Modification) Type	Bioassay Class	ADV Score (kcal mol^{-1})
AHFCA1 (57)	2'-Alkyl	1	-7.46 ± 0.03
AHFCA2 (58)	2'-Alkyl	2	-6.70
AHFCA3 (59)	2'-Alkyl	3	-7.40
AHFCA4 (60)	2'-Alkyl	2	-6.70
AHFCA5 (61)	2'-Alkyl	3	-7.30
HFCA (70)	2'-Alkyl	8	-5.03 ± 0.03
80	2'-Alkyl	8	-5.50
87	Ring condensation	6	-6.87 ± 0.06
106	3'-carboxylic acid	8	-5.27 ± 0.12
120	2'-Alkyl	5	-6.44 ± 0.03
AHFCA7 (121)	2'-Alkyl	3	-6.92 ± 0.03
AHFCA8E (122)	2'-Alkyl	4	-6.26 ± 0.07
123	2'-Alkyl	2	-7.20
124	2'-Alkyl	3	-6.44 ± 0.03
125-*r*	2'-Alkyl	3	-6.97 ± 0.03
125-*s*	2'-Alkyl	3	-6.40
126	2'-ether	8	-5.70
127	2'-ether	6	-5.87 ± 0.11
128	3'-carboxylic acid	5	-6.87 ± 0.06
131	4'-hydroxymethyl	3	-7.50
132	4'-hydroxymethyl	5	-7.20
133	4'-hydroxymethyl	8	-7.40
134	4'-hydroxymethyl	5	-7.50
152	4'-hydroxymethyl	8	-7.40
158	4'-methylthiophene	4	-7.30

Table 8-10 – *n*-, iso- and *tert*-alkyl moieties of AHFCA structures used in AutoDOCK Vina docking to the MmfR:AHFCA2 crystal structure. *x* denotes the length of the alkyl chain. Assigned names are indicated underneath the IUPAC names in parentheses. PDB structures are available in the Electronic Supplementary Materials.

x	2'-*n*-alkyl	2'-isoalkyl	2'-*tert*-alkyl
0	- (Compound 70)	-	-
1	methyl (Compound 80)	-	-
2	ethyl (Compound 120)	methylethyl (isopropyl) (Compound 124)	1,1-dimethylethyl (*tert*-butyl)
3	propyl (AHFCA2)	2-methylpropyl (isobutyl) (AHFCA1)	2,2-dimethylpropyl (*tert*-pentyl)
4	butyl (AHFCA4)	3-methylbutyl (isopentyl) (AHFCA3)	3,3-dimethylbutyl (*tert*-hexyl)
5	pentyl (AHFCA5)	4-methylpentyl (isohexyl) (AHFCA6)	4,4-dimethylpentyl (*tert*-heptyl)
6	hexyl (AHFCA7)	5-methylhexyl (isoyheptyl) (AHFCA8)	5,5-dimethylhexyl (*tert*-octyl)
7	heptyl (AHFCA8E)	6-methylheptyl (isooctyl) (AHFCA10)	6,6-dimethylheptyl (*tert*-nonyl)
8	octyl	7-methyloctyl (isononyl)	7,7-dimethyloctyl (*tert*-decyl)
9	nonyl	8-methylnonyl (isodecyl)	8,8-dimethylnonyl (*tert*-undecyl)
10	decyl	9-methyldecyl (isoundecyl)	9,9-dimethyldecyl (*tert*-dodecyl)
11	undecyl	10-methylundecyl (isododecyl)	10,10-dimethylundecyl (*tert*-tridecyl)
12	docecyl	11-methyldocecyl (isotridecyl)	11,11-dimethyldodecyl (*tert*-tetradecyl)

Table 8-11 – Docking scores of AHFCA structures indicated in Table 8-10 to the crystal structure of MmfR:AHFCA2. Docked AHFCA conformations were selected based on interactions formed with at least Tyr85 and Tyr144 *via* the carboxylic acid moiety, with SEM indicated for 3 replicates. Highest scoring docked structure is shown in **bold**, with arrows showing increasing affinity in the indicated direction. Green indicates $\Delta G \leq \Delta G_{AHFCA2}$, yellow indicates $\Delta G_{AHFCA2} < \Delta G \leq \Delta G_{X=0}$, and red indicates $\Delta G > \Delta G_{X=0}$.

	ΔG (kcal mol^{-1})		
x	R (*n*-alkyl)	R (isoalkyl)	R (*tert*-alkyl)
0	-5.02 ± 0.02	-	-
1	-5.50	-	-
2	-6.44 ± 0.02 →	-6.88 ± 0.02 ←	-5.60
3	-6.70 →	**-7.46 ± 0.02** ←	-5.40
4	-6.70 →	-7.40 ←	-5.96 ± 0.02
5	-7.30 ←	-7.12 ± 0.02 ←	-4.28 ± 0.18
6	-6.92 ± 0.02 ←	-6.46 ± 0.02 ←	-4.30 ± 0.34
7	-6.26 ± 0.05 ←	-5.54 ± 0.20 ←	-2.80 ± 0.61
8	-5.52 ± 0.13 ←	-4.74 ± 0.05 ←	-1.48 ± 0.59
9	-5.20 ± 0.10 ←	-3.38 ± 0.37 ←	-1.02 ± 0.40
10	-3.72 ± 0.34 ←	-2.42 ± 0.41 ←	-1.58 ± 0.35
11	-3.14 ± 0.64 ←	-2.54 ± 0.42 ←	-1.28 ± 0.24
12	-2.32 ± 1.14 ←	-1.76 ± 0.45 ←	-0.06 ± 0.64

Table 8-12 – Docking scores of AHFCA structures indicated in Table 8-10 to the Phyre^2 model of AvaL1 based on the structure of MmfR:AHFCA2. Docked AHFCA conformations were selected based on interactions formed with at least Tyr85 and Tyr144 *via* the carboxylic acid moiety, with SEM indicated for 3 replicates. Highest scoring docked structure is shown in **bold**, with arrows showing increasing affinity in the indicated direction. Green indicates $\Delta G \leq \Delta G_{AHFCA2}$, yellow indicates $\Delta G_{AHFCA2} < \Delta G \leq \Delta G_{X=0}$, and red indicates $\Delta G > \Delta G_{X=0}$.

	ΔG (kcal mol^{-1})		
x	R (*n*-alkyl)	R (isoalkyl)	R (*tert*-alkyl)
0	-4.70	-	-
1	-5.40	-	-
2	-5.22 ± 0.10 →	-5.86 ± 0.02 ←	-5.60
3	-5.26 ± 0.10 →	-5.78 ± 0.02 ←	-4.67 ± 0.23
4	-5.10 ± 0.09 →	-5.54 ± 0.02 →	-5.73 ± 0.03
5	-5.34 ± 0.04 ←	-5.26 ± 0.04 →	-5.27 ± 0.03
6	-5.46 ± 0.05 →	-5.70 ± 0.15 ←	-5.33 ± 0.03
7	-5.60 ± 0.03 ←	-5.58 ± 0.05 ←	-4.73 ± 0.55
8	-5.26 ± 0.06 →	-5.33 ± 0.03 →	**-6.20**
9	-5.40 ± 0.16 ←	-5.00 ± 0.20 ←	-4.73 ± 0.27
10	-5.56 ± 0.06 ←	-5.10 ± 0.21 ←	-4.87 ± 0.09
11	-5.34 ± 0.07 ←	-4.97 ± 0.18 ←	-3.13 ± 0.48
12	-5.16 ± 0.07 ←	-4.40 ± 0.10 →	4.60 ± 0.06

Table 8-13 – Docking scores of AHFCA structures indicated in Table 8-10 to the Phyre^2 model of SclM1 based on the structure of MmfR:AHFCA2. Docked AHFCA conformations were selected based on interactions formed with at least Tyr85 and Tyr144 *via* the carboxylic acid moiety, with SEM indicated for 3 replicates. Highest scoring docked structure is shown in **bold**, with arrows showing increasing affinity in the indicated direction. **Green** indicates $\Delta G \leq \Delta G_{AHFCA2}$, **yellow** indicates $\Delta G_{AHFCA2} < \Delta G \leq \Delta G_{X=0}$, and **red** indicates $\Delta G > \Delta G_{X=0}$.

	ΔG (kcal mol^{-1})		
x	R (*n*-alkyl)	R (isoalkyl)	R (*tert*-alkyl)
0	-4.97 ± 0.03	-	-
1	-5.20	-	-
2	-5.90	-6.23 ± 0.13	-5.47 ± 0.08
3	-6.50	**-7.27 ± 0.03**	-3.77 ± 0.23
4	-5.75 ± 0.03	-6.80	-5.17 ± 0.03
5	-6.40 ± 0.06	-4.77 ± 0.03	-3.53 ±0.12
6	-5.50 ± 0.21	-5.03 ± 0.35	-2.87 ± 0.77
7	-5.40 ± 0.10	-4.47 ± 0.12	-1.63 ± 0.37
8	-5.07 ± 0.03	-4.77 ± 0.17	-2.73 ± 0.09
9	-5.07 ± 0.07	-4.47 ± 0.09	-2.87 ± 0.23
10	-4.77 ± 0.03	-3.77 ± 0.09	-2.10 ± 0.12
11	-4.43 ± 0.07	-3.13 ± 0.35	-1.00 ± 0.38
12	-4.23 ± 0.09	-2.47 ± 0.38	-2.27 ± 0.20

Table 8-14 – Docking scores of AHFCA structures indicated in Table 8-10 to the Phyre^2 model of SgnR based on the structure of MmfR:AHFCA2. Docked AHFCA conformations were selected based on interactions formed with at least Tyr85 and Tyr144 *via* the carboxylic acid moiety, with SEM indicated for 3 replicates. Highest scoring docked structure is shown in **bold**, with arrows showing increasing affinity in the indicated direction. **Green** indicates $\Delta G \leq \Delta G_{AHFCA2}$, **yellow** indicates $\Delta G_{AHFCA2} < \Delta G \leq \Delta G_{X=0}$, and **red** indicates $\Delta G > \Delta G_{X=0}$.

	ΔG (kcal mol^{-1})		
x	R (*n*-alkyl)	R (isoalkyl)	R (*tert*-alkyl)
0	-4.80	-	-
1	-5.20	-	-
2	-5.27 ± 0.03	-5.80	-5.63 ± 0.42
3	-6.20	**-6.93 ± 0.03**	-3.77 ± 0.03
4	-5.53 ± 0.03	-6.60	-4.50 ± 0.12
5	-6.60 ± 0.06	-6.17 ± 0.07	-3.77 ± 0.17
6	-6.00 ±0.05	-5.53 ± 0.03	-3.03 ± 0.12
7	-5.47 ±0.03	-4.93 ± 0.09	-2.97 ± 0.09
8	-5.33 ±0.07	-5.37 ± 0.09	-3.10 ± 0.40
9	-5.60	-4.87 ± 0.07	-3.07 ± 0.19
10	-5.27 ±0.03	-4.63 ± 0.03	-2.97 ± 0.09
11	-4.93 ± 0.13	-3.70 ± 0.06	-2.23 ± 0.03
12	-4.47 ±0.24	-3.37 ± 0.18	-1.13 ± 0.73

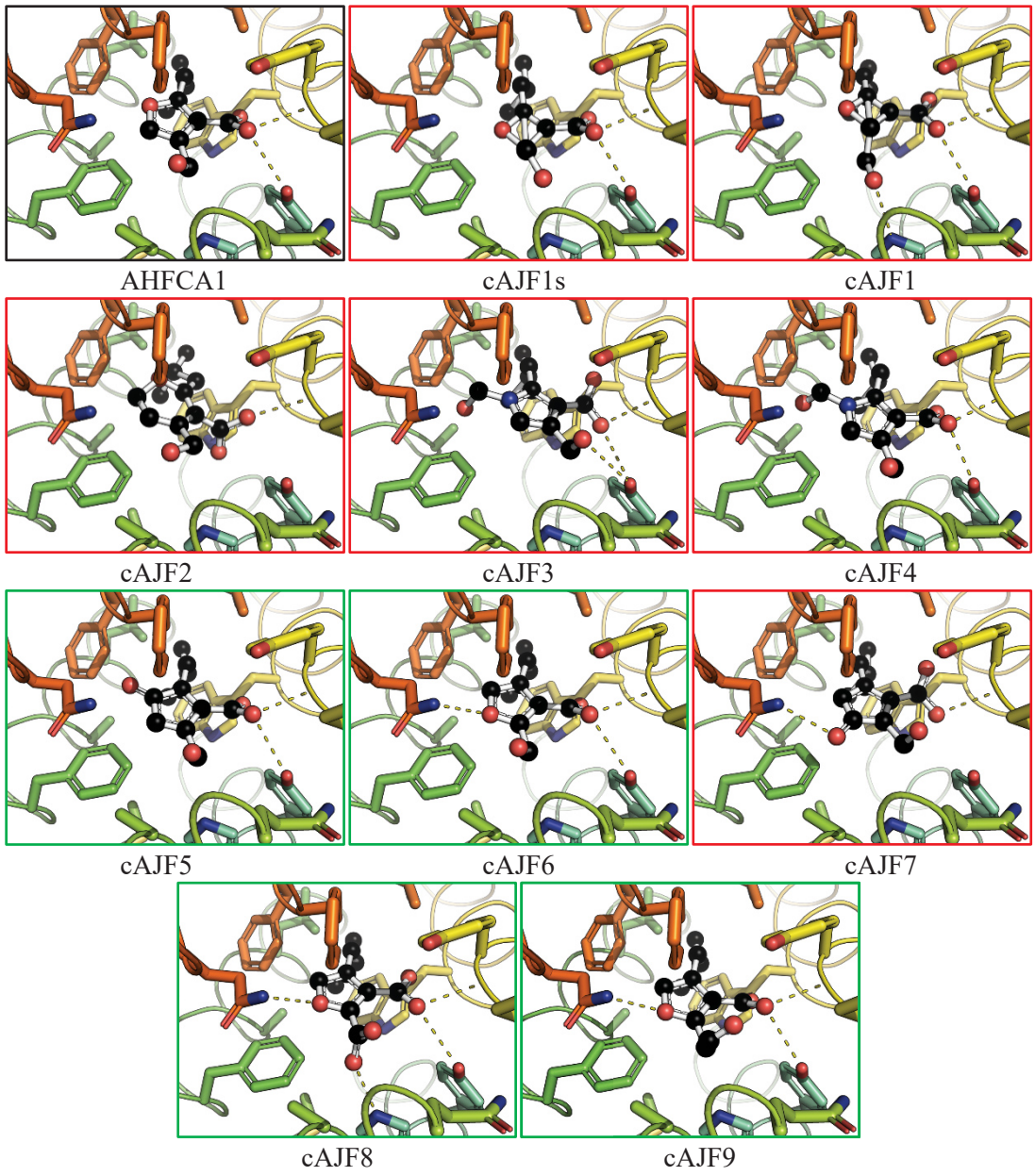


Figure 8-20 – ADV docking conformations of the structures listed in Figure 6-11 to the MmfR:AHHFCA2 crystal structure. Molecules scoring better than AHFCA are enclosed in a **green** box, and those scoring worse in a **red** box.

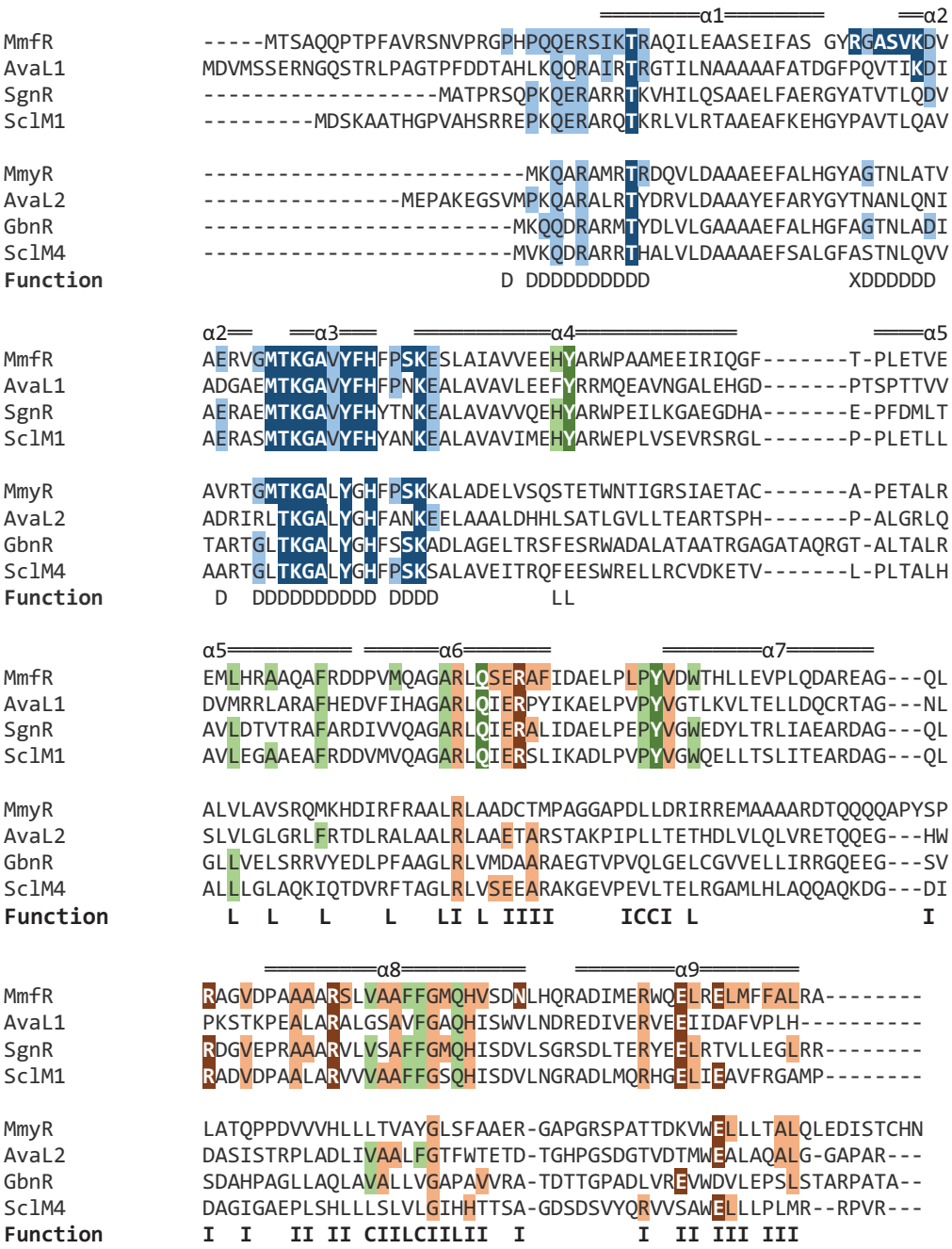


Figure 8-21 – MUSCLE alignment of AHFCA and pseudo-AHFCA receptors annotated with functionally important residues and their conservation with MmfR. Functional residues of each TFR are indicated by their colour; **blue** indicates predicted DNA-interacting residues (Chapter 5.3.1), with stronger predictions in a darker shade. **Brown** indicates the dimer interface, with darker orange denoting electrostatically interacting residues. **Green** residues indicate the ligand-binding pocket, with darker residues interacting directly with AHFCAs. Where residues are conserved across orthologues, colour-coding is maintained vertically. Helices are denoted with the character ═. Comments. Function is also indicated underneath the protein sequences; D = residue involved in DNA interface, L = residues found in the LBP, and I = interface residues. X = D + I, and C = I + L.

8.6 Appendix F

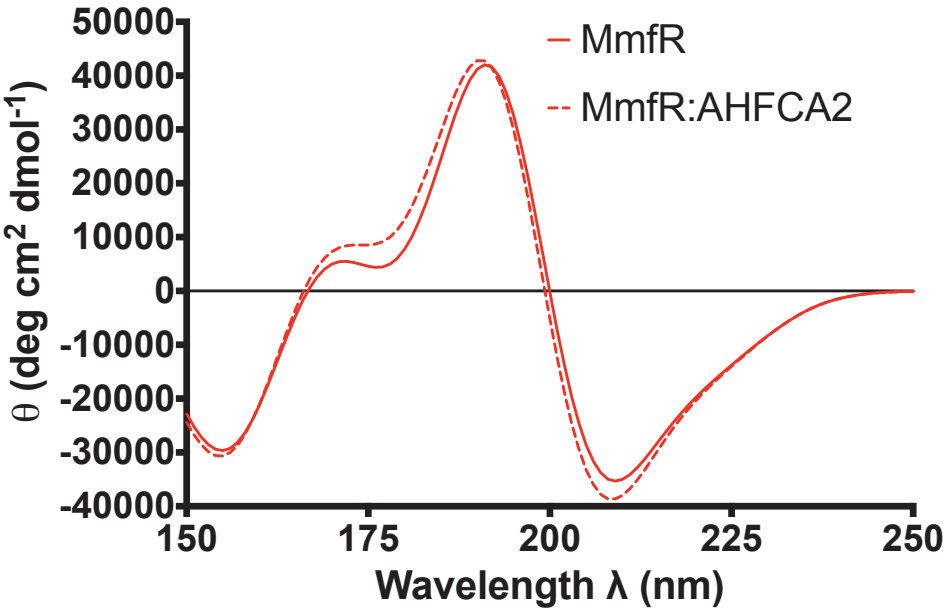


Figure 8-22 – Theoretical 150-250 nm circular dichroism spectra of MmfR and MmfR:AHFCA2 calculated using DichroCalc based on 1.5 Å crystal structures.

9 Bibliography

1. Dharmaraj S. Marine *Streptomyces* as a novel source of bioactive substances. World Journal of Microbiology & Biotechnology. 2010 Dec;26(12):2123-39.
2. Bignell DRD, Fyans JK, Cheng Z. Phytotoxins produced by plant pathogenic *Streptomyces* species. Journal of Applied Microbiology. 2014 Feb;116(2):223-35.
3. Kritzman G, ShaniCahani A, Kirshner B, Riven Y, Bar Z, Katan J, et al. Pod wart disease of peanuts. Phytoparasitica. 1996;24(4):293-304.
4. King RR, Calhoun LA. The thaxtomin phytotoxins: Sources, synthesis, biosynthesis, biotransformation and biological activity. Phytochemistry. 2009 May;70(7):833-41.
5. Zhang YC, Bignell DRD, Zuo R, Fan QR, Huguet-Tapia JC, Ding YS, et al. Promiscuous Pathogenicity Islands and Phylogeny of Pathogenic *Streptomyces* spp. Molecular Plant-Microbe Interactions. 2016 Aug;29(8):640-50.
6. Hotson IK. The avermectins: A new family of antiparasitic agents. Journal of the South African Veterinary Association-Tydskrif Van Die Suid-Afrikaanse Veterinere Vereniging. 1982;53(2):87-90.
7. Halley BA, Vandenheuvel WJA, Wislocki PG. Environmental effects of the usage of avermectins in livestock. Veterinary Parasitology. 1993 Jun;48(1-4):109-25.
8. Bloom RA, Matheson JC. Environmental assessment of avermectins by the US Food and Drug Administration. Veterinary Parasitology. 1993 Jun;48(1-4):281-94.
9. Wang WJ, Lu HQ, Zhong P, Sun GY. Nystagmus following acute avermectins poisoning. Toxicology and Industrial Health. 2016 Apr;32(4):751-2.
10. Kapadia M, Rolston KVI, Han XY. Invasive *Streptomyces* infections: Six cases and literature review. American Journal of Clinical Pathology. 2007 Apr;127(4):619-24.
11. Borelli D, Middelveen M. Actinomycetoma caused by S*treptomyces somaliensis*. Archives of Dermatology. 1986 Oct;122(10):1097-8.
12. Quintana ET, Wierzbicka K, Mackiewicz P, Osman A, Fahal AH, Hamid ME, et al. *Streptomyces sudanensis* sp. nov., a new pathogen isolated from patients with actinomycetoma. Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology. 2008 Mar;93(3):305-13.
13. *Streptomyces coelicolor* Exploring The Invisible2013 [cited 28/8/18]. Just a straight foward photograph of the common soil bacterium S*treptomyces coelicolor* , reavealing its inherent complexity and beauty. These bacteria are notable because they have a highly complex secondary metabolism through which they many clinically useful antibiotics and compounds. They are the microbial world's chemists.]. Available from: https://exploringtheinvisible.com/2013/04/16/streptomyces-coelicolor/.
14. Zacharia VM, Traxler MF. Bacteria: Exploring new horizons. Elife. 2017 Jan;6:3.
15. Challis GL, Hopwood DA. Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by Streptomyces species. Proceedings of the National Academy of Sciences of the United States of America. 2003;100 Suppl 2(Suppl 2):14555-61.
16. Yim G, Wang HH, Davies J. Antibiotics as signalling molecules. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2007;362(1483):1195-200.
17. Jauri PV, Bakker MG, Salomon CE, Kinkel LL. Subinhibitory Antibiotic Concentrations Mediate Nutrient Use and Competition among Soil Streptomyces. Plos One. 2013 Dec;8(12):6.
18. Abrudan MI, Smakman F, Grimbergen AJ, Westhoff S, Miller EL, van Wezel GP, et al. Socially mediated induction and suppression of antibiosis during bacterial coexistence. Proceedings of the National Academy of Sciences of the United States of America. 2015 Sep;112(35):11054-9.
19. Seipke RF, Kaltenpoth M, Hutchings MI. Streptomyces as symbionts: an emerging and widespread theme? FEMS Microbiology Reviews. 2012;36(4):862-76.
20. Rey T, Dumas B. Plenty Is No Plague: Streptomyces Symbiosis with Crops. Trends in Plant Science. 2017 Jan;22(1):30-7.
21. Ceapa CD, Vazquez-Hernandez M, Rodriguez-Luna SD, Vazquez APC, Suarez VJ, Rodriguez-Sanoja R, et al. Genome mining of Streptomyces scabrisporus NF3 reveals symbiotic features including genes related to plant interactions. Plos One. 2018 Feb;13(2):27.
22. Vurukonda S, Giovanardi D, Stefani E. Plant Growth Promoting and Biocontrol Activity of Streptomyces spp. as Endophytes. International Journal of Molecular Sciences. 2018 Apr;19(4):26.
23. Thilagam R, Hemalatha N. Plant growth promotion and chilli anthracnose disease suppression ability of rhizosphere soil actinobacteria. Journal of Applied Microbiology. 2019 Jun;126(6):1835-49.
24. Li ZY. Advances in Marine Microbial Symbionts in the China Sea and Related Pharmaceutical Metabolites. Marine Drugs. 2009 Jun;7(2):113-29.
25. G Nair A, Dharmaraj S, Dhevendaran K. Occurrence of Sponges Associated Streptomyces and its Antimicrobial Activity2011.
26. Brana AF, Fiedler HP, Nava H, Gonzalez V, Sarmiento-Vizcaino A, Molina A, et al. Two Streptomyces Species Producing Antibiotic, Antitumor, and Anti-Inflammatory Compounds Are Widespread Among Intertidal Macroalgae and Deep-Sea Coral Reef Invertebrates from the Central Cantabrian Sea. Microbial Ecology. 2015 Apr;69(3):512-24.
27. Van Arnam EB, Currie CR, Clardy J. Defense contracts: molecular protection in insect-microbe symbioses. Chemical Society Reviews. 2018 Mar;47(5):1638-51.

28. Chevrette MG, Carlson CM, Ortega HE, Thomas C, Ananiev GE, Barns KJ, et al. The antimicrobial potential of Streptomyces from insect microbiomes. Nature Communications. 2019 Jan;10:11.
29. Liu HS, Chen ZB, Zhu GL, Wang LP, Du YQ, Wang Y, et al. Phenolic polyketides from the marine alga-derived Streptomyces sp OUCMDZ-3434. Tetrahedron. 2017 Sep;73(36):5451-5.
30. Calcott MJ, Ackerley DF, Knight A, Keyzers RA, Owen JG. Secondary metabolism in the lichen symbiosis. Chemical Society Reviews. 2018 Mar;47(5):1730-60.
31. de Lima Procopio RE, da Silva IR, Martins MK, de Azevedo JL, de Araujo JM. Antibiotics produced by *Streptomyces*. Brazilian Journal of Infectious Diseases. 2012 Sep-Oct;16(5):466-71.
32. Schatz A, Bugie E, Waksman S. Streptomycin, a substance exhibiting antibiotic activity against Gram-positive and Gram-negative bacteria. *Proc Exp Biol Med*. 1944 January;55:66-9.
33. Waksman SA, Woodruff HB. Bacteriostatic and Bacteriocidal Substances produced by a soil *Actinomyces*. Proceedings of the Society for Experimental Biology and Medicine. 1940 (45):609.
34. Deak D, Outterson K, Powers JH, Kesselheim AS. Progress in the Fight Against Multidrug-Resistant Bacteria? A Review of US Food and Drug Administration-Approved Antibiotics, 2010-2015. Annals of Internal Medicine. 2016 Sep;165(5):363-+.
35. Sherpa RT, Reese CJ, Aliabadi HM. Application of iChip to Grow "Uncultivable" Microorganisms and its Impact on Antibiotic Discovery. Journal of Pharmacy and Pharmaceutical Sciences. 2015;18(3):303-15.
36. Conn J. The Pigment Production of A*ctinomyces coelicolor* and *A. violaceus-ruber*. Journal of Bacteriology. 1943;46:133-49.
37. Kutzner H, SA. *Streptomyces coelicolor* Müller and S*treptomyces violaceoruber* Waksman and Curtis, two distinctly different organisms. J Bacteriol. 1959 October;78(4):528-38.
38. Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, et al. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature. 2002 May;417(6885):141-7.
39. Han K, Li ZF, Peng R, Zhu LP, Zhou T, Wang LG, et al. Extraordinary expansion of a *Sorangium cellulosum* genome from an alkaline milieu. Scientific Reports. 2013 Jul;3:7.
40. Moran NA, Bennett GM. Small, Smaller, Smallest: The Origins and Evolution of Ancient Dual Symbioses in a Phloem-Feeding Insect. Genome Biology and Evolution. 2013;5(9):1675-88.
41. Haug I, Weissenborn A, Brolle D, Bentley S, Kieser T, Altenbuchner J. Streptomyces coelicolor A3(2) plasmid SCP2*: deductions from the complete sequence. Microbiology. 2003;149(2):505-13.
42. Bentley SD, Brown S, Murphy LD, Harris DE, Quail MA, Parkhill J, et al. SCP1, a 356,023 bp linear plasmid adapted to the ecology and developmental biology of its host, *Streptomyces coelicolor* A3(2). Molecular Microbiology. 2004 Mar;51(6):1615-28.
43. Kirby R, Wright LF, Hopwood DA. Plasmid-determined antibiotic synthesis and resistance in *Streptomyces coelicolor*. Nature. 1975;254(5497):265-7.
44. Burg RW, Miller BM, Baker EE, Birnbaum J, Currie SA, Hartman R, et al. Avermectins, new family of potent anthelmintic agents: producing organism and fermentation. Antimicrobial Agents and Chemotherapy. 1979;15(3):361-7.
45. Omura S, Ikeda H, Ishikawa J, Hanamoto A, Takahashi C, Shinose M, et al. Genome sequence of an industrial microorganism *Streptomyces avermitilis*: Deducing the ability of producing secondary metabolites. Proceedings of the National Academy of Sciences of the United States of America. 2001 Oct;98(21):12215-20.
46. Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, et al. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. Nature Biotechnology. 2003 May;21(5):526-31.
47. Zerikly M, Challis GL. Strategies for the Discovery of New Natural Products by Genome Mining. Chembiochem. 2009 Mar;10(4):625-33.
48. Lawrence JG, Roth JR. Selfish operons: horizontal transfer may drive the evolution of gene clusters. Genetics. 1996;143(4):1843-60.
49. Ballouz S, Francis AR, Lan R, Tanaka MM. Conditions for the Evolution of Gene Clusters in Bacterial Genomes. PLOS Computational Biology. 2010;6(2):e1000672.
50. Liu G, Chater KF, Chandra G, Niu GQ, Tan HR. Molecular Regulation of Antibiotic Biosynthesis in *Streptomyces*. Microbiology and Molecular Biology Reviews. 2013 Mar;77(1):112-43.
51. Tyurin AP, Alferova VA, Korshun VA. Chemical Elicitors of Antibiotic Biosynthesis in Actinomycetes. Microorganisms. 2018;6(2):52.
52. Bibb MJ. Regulation of secondary metabolism in *Streptomycetes*. Current Opinion in Microbiology. 2005 Apr;8(2):208-15.
53. van Wezel GP, McDowall KJ. The regulation of the secondary metabolism of *Streptomyces*: new links and experimental advances. Natural Product Reports. 2011;28(7):1311-33.
54. Wright LF, Hopwood DA. Actinorhodin is a chromosomally-determined antibiotic in S*treptomyces coelicolor A*3(2). Journal of General Microbiology. 1976;96(OCT):289-97.
55. Rudd BAM, Hopwood DA. Genetics of actinorhodin biosynthesis by *Streptomyces coelicolor* A3(2). Journal of General Microbiology. 1979;114(SEP):35-43.
56. Rudd BAM, Hopwood DA. A pigmented mycelial antibiotic in *Streptomyces coelicolor*: control by a chromosomal gene cluster. Journal of General Microbiology. 1980;119(AUG):333-40.

57. Pawlik K, Kotowska M, Chater KF, Kuczek K, Takano E. A cryptic type I polyketide synthase (*cpk*) gene cluster in *Streptomyces coelicolor* A3(2). Archives of Microbiology. 2007 Feb;187(2):87-99.
58. Gomez-Escribano JP, Song L, Fox DJ, Yeo V, Bibb MJ, Challis GL. Structure and biosynthesis of the unusual polyketide alkaloid coelimycin P1, a metabolic product of the *cpk* gene cluster of *Streptomyces coelicolor* M145. Chemical Science. 2012 2012;3(9):2716-20.
59. Hopwood DA, Wright HM. CDA is a New Chromosomally-determined Antibiotic from *Streptomyces*
coelicolor A3(2) Journal of General Microbiology. 1983;129(DEC):3575-9.
60. Blin K, Wolf T, Chevrette MG, Lu XW, Schwalen CJ, Kautsar SA, et al. antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. Nucleic Acids Research. 2017 Jul;45(W1):W36-W41.
61. Fernandezmoreno MA, Caballero JL, Hopwood DA, Malpartida F. The a*ct* cluster contains regulatory and antibiotic export genes, direct targets for translational control by the *bldA* transfer-RNA gene of S*treptomyces*. Cell. 1991 Aug;66(4):769-80.
62. Hackl S, Bechthold A. The Gene *bldA*, a Regulator of Morphological Differentiation and Antibiotic Production in *Streptomyces*. Archiv Der Pharmazie. 2015 Jul;348(7):455-62.
63. Tahlan K, Yu Z, Xu Y, Davidson AR, Nodwell JR. Ligand Recognition by ActR, a TetR-Like Regulator of Actinorhodin Export. Journal of Molecular Biology. 2008 Nov;383(4):753-61.
64. Cuthbertson L, Nodwell JR. The TetR Family of Regulators. Microbiology and Molecular Biology Reviews. 2013 Sep;77(3):440-75.
65. Ramos JL, Martínez-Bueno M, Molina-Henares AJ, Terán W, Watanabe K, Zhang X, et al. The TetR family of transcriptional repressors. Microbiology and molecular biology reviews : MMBR. 2005;69(2):326-56.
66. Takano E, Kinoshita H, Mersinias V, Bucca G, Hotchkiss G, Nihira T, et al. A bacterial hormone (the SCB1) directly controls the expression of a pathway-specific regulatory gene in the cryptic type I polyketide biosynthetic gene cluster of *Streptomyces coelicolor*. Molecular Microbiology. 2005 Apr;56(2):465-79.
67. O'Rourke S, Wietzorrek A, Fowler K, Corre C, Challis GL, Chater KF. Extracellular signalling, translational control, two repressors and an activator all contribute to the regulation of methylenomycin production in S*treptomyces coelicolor*. Molecular Microbiology. 2009 Feb;71(3):763-78.
68. Onaka H, Nakagawa T, Horinouchi S. Involvement of two A-factor receptor homologues in *Streptomyces coelicolor* A3(2) in the regulation of secondary metabolism and morphogenesis. Molecular Microbiology. 1998 May;28(4):743-53.
69. Bhukya H, Bhujbalrao R, Bitra A, Anand R. Structural and functional basis of transcriptional regulation by TetR family protein CprB from *S. coelicolor* A3(2). Nucleic acids research. 2014 2014 Nov 1 (Epub 2014 Aug;42(15):10122-33.
70. Xu DL, Seghezzi N, Esnault C, Virolle MJ. Repression of Antibiotic Production and Sporulation in *Streptomyces coelicolor* by Overexpression of a TetR Family Transcriptional Regulator. Applied and Environmental Microbiology. 2010 Dec;76(23):7741-53.
71. Uguru GC, Stephens KE, Stead JA, Towle JE, Baumberg S, McDowall KJ. Transcriptional activation of the pathway-specific regulator of the actinorhodin biosynthetic genes in *Streptomyces coelicolor*. Molecular Microbiology. 2005 Oct;58(1):131-50.
72. Fu JF, Zong GL, Zhang PP, Zhao ZL, Ma JX, Pang XH, et al. XdhR negatively regulates actinorhodin biosynthesis in *Streptomyces coelicolor* M145. Fems Microbiology Letters. 2017 Nov;364(22):9.
73. Lee HN, Huang JQ, Im JH, Kim SH, Noh JH, Cohen SN, et al. Putative TetR Family Transcriptional Regulator SCO1712 Encodes an Antibiotic Downregulator in *Streptomyces coelicolor*. Applied and Environmental Microbiology. 2010 May;76(9):3039-43.
74. Ou XJ, Zhang B, Zhang L, Zhao GP, Ding XM. Characterization of *rrdA*, a TetR Family Protein Gene Involved in the Regulation of Secondary Metabolism in *Streptomyces coelicolor*. Applied and Environmental Microbiology. 2009 Apr;75(7):2158-65.
75. Sivapragasam S, Grove A. *Streptomyces coelicolor* XdhR is a direct target of (p)ppGpp that controls expression of genes encoding xanthine dehydrogenase to promote purine salvage. Molecular Microbiology. 2016 May;100(4):701-18.
76. Li X, Wang J, Li SS, Ji JJ, Wang WS, Yang KQ. ScbR- and ScbR2-mediated signal transduction networks coordinate complex physiological responses in *Streptomyces coelicolor*. Scientific Reports. 2015 Oct;5:12.
77. Zhu JY, Sun D, Liu WS, Chen Z, Li JL, Wen Y. AvaR2, a pseudo-butyrolactone receptor homologue from *Streptomyces avermitilis*, is a pleiotropic repressor of avermectin and avenolide biosynthesis and cell growth. Molecular Microbiology. 2016 Nov;102(4):562-78.
78. Zhu JY, Chen Z, Li JL, Wen Y. AvaR1, a Butenolide-Type Autoregulator Receptor in *Streptomyces avermitilis*, Directly Represses Avenolide and Avermectin Biosynthesis and Multiple Physiological Responses. Frontiers in Microbiology. 2017 Dec;8:15.
79. He HR, Ye L, Li C, Wang HY, Guo XW, Wang XJ, et al. SbbR/SbbA, an Important ArpA/AfsA-Like System, Regulates Milbemycin Production in Streptomyces bingchenggensis. Frontiers in Microbiology. 2018 May;9:17.

80. Yamada Y. Butyrolactone Autoregulators, Inducers of Secondary Metabolites, in *Streptomyces*. Actinomycetologica. 1995;9(1):57-65.
81. Onaka H, Ando N, Nihira T, Yamada Y, Beppu T, Horinouchi S. Cloning and characterization of the A-factor receptor gene from *Streptomyces griseus*. Journal of Bacteriology. 1995 Nov;177(21):6083-92.
82. Ohnishi Y, Yamazaki H, Kato JY, Tomono A, Horinouchi S. AdpA, a central transcriptional regulator in the A-factor regulatory cascade that leads to morphological development and secondary metabolism in *Streptomyces griseus*. Bioscience Biotechnology and Biochemistry. 2005 Mar;69(3):431-9.
83. Ohnishi Y, Ishikawa J, Hara H, Suzuki H, Ikenoya M, Ikeda H, et al. Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. Journal of Bacteriology. 2008 Jun;190(11):4050-60.
84. Ohnuki T, Imanaka T, Aiba S. Self-cloning in *Streptomyces griseus* of an *str* gene cluster for streptomycin biosynthesis and streptomycin resistance. Journal of Bacteriology. 1985;164(1):85-94.
85. Distler J, Braun C, Ebert A, Piepersberg W. Gene cluster for streptomycin biosynthesis in *Streptomyces griseus*: Analysis of a central region including the major resistance gene. Molecular & General Genetics. 1987 Jun;208(1-2):204-10.
86. Distler J, Ebert A, Mansouri K, Pissowotzki K, Stockmann M, Piepersberg W. Gene cluster for streptomycin biosynthesis in S*treptomyces griseus:* Nucleotide-sequence of 3 genes and analysis of transcriptional activity. Nucleic Acids Research. 1987 Oct;15(19):8041-56.
87. Khokhlov AS, Anisova LN, Tovarova, II, Kleiner EM, Kovalenko IV, Krasilnikova OI, et al. Effect of A-factor on the growth of asporogenous mutants of *Streptomyces griseus*, not producing this factor. Zeitschrift Fur Allgemeine Mikrobiologie. 1973;13(8):647-55.
88. Miyake K, Horinouchi S, Yoshida M, Chiba N, Mori K, Nogawa N, et al. Detection and properties of A-factor-binding protein from *Streptomyces griseus*. Journal of Bacteriology. 1989 Aug;171(8):4298-302.
89. Vujaklija D, Ueda K, Hong SK, Beppu T, Horinouchi S. Identification of an A-factor-dependent promoter in the streptomycin biosynthetic gene cluster of S*treptomyces griseus*. Molecular & General Genetics. 1991 Sep;229(1):119-28.
90. Vujaklija D, Horinouchi S, Beppu T. Detection of an A-factor-responsive protein that binds to the upstream activation sequence of s*trR,* a regulatory gene for streptomycin biosynthesis in S*treptomyces griseus*. Journal of Bacteriology. 1993 May;175(9):2652-61.
91. Ohnishi Y, Kameyama S, Onaka H, Horinouchi S. The A-factor regulatory cascade leading to streptomycin biosynthesis in *Streptomyces griseus*: identification of a target gene of the A-factor receptor. Molecular Microbiology. 1999 Oct;34(1):102-11.
92. Arakawa K, Tsuda N, Taniguchi A, Kinashi H. The Butenolide Signaling Molecules SRB1 and SRB2 Induce Lankacidin and Lankamycin Production in *Streptomyces rochei*. Chembiochem. 2012 Jul;13(10):1447-57.
93. Kitani S, Miyamoto KT, Takamatsu S, Herawati E, Iguchi H, Nishitomi K, et al. Avenolide, a *Streptomyces* hormone controlling antibiotic production in *Streptomyces avermitilis*. Proceedings of the National Academy of Sciences of the United States of America. 2011 Sep;108(39):16410-5.
94. Idowu G. Investigations into the biosynthesis and mode of action of methylenomycin antibiotics from *Streptomyces coelicolor* [PhD]. WRAP: University of Warwick; 2017.
95. Corre C, Song L, O'Rourke S, Chater KF, Challis GL. 2-Alkyl-4-hydroxymethylfuran-3-carboxylic acids, antibiotic production inducers discovered by *Streptomyces coelicolor* genome mining. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(45):17510-5.
96. Yamasaki M, Ikuto Y, Ohira A, Chater K, Kinashi H. Limited regions of homology between linear and circular plasmids encoding methylenomycin biosynthesis in two independently isolated S*treptomycetes*. Microbiology-Sgm. 2003 May;149:1351-6.
97. Xu GM, Wang JA, Wang LQ, Tian XY, Yang HH, Fan KQ, et al. "Pseudo" Gamma-Butyrolactone Receptors Respond to Antibiotic Signals to Coordinate Antibiotic Biosynthesis. Journal of Biological Chemistry. 2010 Aug;285(35):27440-8.
98. Takano E, Chakraburtty R, Nihira T, Yamada Y, Bibb MJ. A complex role for the gamma-butyrolactone SCB1 in regulating antibiotic production in *Streptomyces coelicolor* A3(2). Molecular Microbiology. 2001 Sep;41(5):1015-28.
99. Wang J, Wang WS, Wang LQ, Zhang GF, Fan KQ, Tan HR, et al. A novel role of 'pseudo' gamma-butyrolactone receptors in controlling gamma-butyrolactone biosynthesis in *Streptomyces*. Molecular Microbiology. 2011 Oct;82(1):236-50.
100. Wang W, Ji J, Li X, Wang J, Li S, Pan G, et al. Angucyclines as signals modulate the behaviors of *Streptomyces coelicolor*. Proceedings of the National Academy of Sciences of the United States of America. 2014 Apr 15;111(15):5688-93.
101. Horinouchi S, Suzuki H, Nishiyama M, Beppu T. Nucleotide Sequence and Transcriptional Analysis of the *Streptomyces griseus* Gene (*afsA*) Responsible for A-Factor Biosynthesis. Journal of Bacteriology. 1989 Feb;171(2):1206-10.
102. D'Alia D, Eggle D, Nieselt K, Hu WS, Breitling R, Takano E. Deletion of the signalling molecule synthase ScbA has pleiotropic effects on secondary metabolite biosynthesis, morphological

differentiation and primary metabolism in *Streptomyces coelicolor* A3(2). Microbial Biotechnology. 2011 Mar;4(2):239-51.
103. Zou ZZ, Du DY, Zhang YY, Zhang JH, Niu GQ, Tan HR. A gamma-butyrolactone-sensing activator/repressor, JadR3, controls a regulatory mini-network for jadomycin biosynthesis. Molecular Microbiology. 2014 Nov;94(3):490-505.
104. Yamada Y, Sugamura K, Kondo K, Yanagimoto M, Okada H. The structure of inducing factors for virginiamycin production in S*treptomyces virginiae*. Journal of Antibiotics. 1987 Apr;40(4):496-504.
105. Kawachi R, Akashi T, Kamitani Y, Sy A, Wangchaisoonthorn U, Nihira T, et al. Identification of an AfsA homologue (BarX) from *Streptomyces virginiae* as a pleiotropic regulator controlling autoregulator biosynthesis, virginiamycin biosynthesis and virginiamycin M-1 resistance. Molecular Microbiology. 2000 Apr;36(2):302-13.
106. Sato K, Nihira T, Sakuda S, Yanagimoto M, Yamada Y. Isolation and structure of a new butyrolactone autoregulator from S*treptomyces* sp. FRI-5. Journal of Fermentation and Bioengineering. 1989;68(3):170-3.
107. Kitani S, Iida A, Izumi TA, Maeda A, Yamada Y, Nihira T. Identification of genes involved in the butyrolactone autoregulator cascade that modulates secondary metabolism in *Streptomyces lavendulae* FRI-5. Gene. 2008 Dec;425(1-2):9-16.
108. Hoshino S, Wakimoto T, Onaka H, Abe I. Chojalactones A-C, Cytotoxic Butanolides Isolated from *Streptomyces* sp. Cultivated with Mycolic Acid Containing Bacterium. Organic Letters. 2015 Mar;17(6):1501-4.
109. Schiewe HJ, Zeeck A. Cineromycins, gamma-butyrolactones and ansamycins by analysis of the secondary metabolite pattern created by a single strain of *Streptomyces*. Journal of Antibiotics. 1999 Jul;52(7):635-42.
110. Sidda JD, Song LJ, Poon V, Al-Bassam M, Lazos O, Buttner MJ, et al. Discovery of a family of gamma-aminobutyrate ureas via rational derepression of a silent bacterial gene cluster. Chemical Science. 2014;5(1):86-9.
111. Sidda J. Discovery and Biosynthesis of Novel Natural Products from *Streptomyces venezuelae* [PhD]. WRAP: University of Warwick; 2015.
112. Poon V. Analysis and exploitation of AHFCA-dependent signalling systems in *Streptomyces* bacteria [PhD]: University of Warwick; 2015.
113. Alberti F, Leng D, Wilkening I, Song L, Tosin M, Corre C. Triggering the expression of a silent gene cluster from genetically intractable bacteria results in scleric acid discovery. Chemical Science. 2019;10(2):453-63.
114. Bignell DRD, Bate N, Cundliffe E. Regulation of tylosin production: role of a TylP-interactive ligand. Molecular Microbiology. 2007 Feb;63(3):838-47.
115. Mutenko H, Makitrinskyy R, Tsypik O, Walker S, Ostash B, Fedorenko V. Genes for biosynthesis of butenolide-like signalling molecules in *Streptomyces ghanaensis*, their role in moenomycin production. Russian Journal of Genetics. 2014 Jun;50(6):563-8.
116. Salehi-Najafabadi Z, Barreiro C, Rodriguez-Garcia A, Cruz A, Lopez GE, Martin JF. The gamma-butyrolactone receptors BulR1 and BulR2 of *Streptomyces tsukubaensis*: tacrolimus (FK506) and butyrolactone synthetases production control. Applied Microbiology and Biotechnology. 2014 Jun;98(11):4919-36.
117. Healy FG, Eaton KP, Limsirichai P, Aldrich JF, Plowman AK, King RR. Characterization of gamma-Butyrolactone Autoregulatory Signaling Gene Homologs in the Angucyclinone Polyketide WS5995B Producer *Streptomyces acidiscabies*. Journal of Bacteriology. 2009 Aug;191(15):4786-97.
118. Meng XX, Wang WZ, Xie ZJ, Li PW, Li Y, Guo ZY, et al. Neomycin biosynthesis is regulated positively by AfsA-g and NeoR in *Streptomyces fradiae* CGMCC 4.7387. Science China-Life Sciences. 2017 Sep;60(9):980-91.
119. Intra B, Euanorasetr J, Nihira T, Panbangred W. Characterization of a gamma-butyrolactone synthetase gene homologue (*stcA*) involved in bafilomycin production and aerial mycelium formation in *Streptomyces* sp. SBI034. Applied Microbiology and Biotechnology. 2016 Mar;100(6):2749-60.
120. Tan GY, Bai LQ, Zhong JJ. Exogenous 1,4-Butyrolactone Stimulates A-Factor-Like Cascade and Validamycin Biosynthesis in *Streptomyces hygroscopicus* 5008. Biotechnology and Bioengineering. 2013 Nov;110(11):2984-93.
121. Gottelt M, Kol S, Gomez-Escribano JP, Bibb M, Takano E. Deletion of a regulatory gene within the *cpk* gene cluster reveals novel antibacterial activity in *Streptomyces coelicolor* A3(2). Microbiology-Sgm. 2010 Aug;156:2343-53.
122. Ehrlich J, Bartz QR, Smith RM, Joslyn DA, Burkholder PR. Chloromycetin, a New Antibiotic From a Soil Actinomycete. Science. 1947 Oct 31;106(2757):417.
123. Ehrlich J, Gottlieb D, Burkholder PR, Anderson LE, Pridham TG. *Streptomyces venezuelae*, N. Sp., the Source of Chloromycetin. J Bacteriol. 1948 Oct;56(4):467-77.
124. Pullan ST, Chandra G, Bibb MJ, Merrick M. Genome-wide analysis of the role of GlnR in *Streptomyces venezuelae* provides new insights into global nitrogen regulation in actinomycetes. Bmc Genomics. 2011 Apr;12:14.
125. Hayashi M, Saito Y, Kawashima S. Calpain activation is essential for membrane fusion of erythrocytes in the presence of exogenous Ca^{2+}. Biochemical and Biophysical Research Communications. 1992 Jan;182(2):939-46.

126. Schellenberg B, Ramel C, Dudler R. *Pseudomonas syringae* Virulence Factor Syringolin A Counteracts Stomatal Immunity by Proteasome Inhibition. Molecular Plant-Microbe Interactions. 2010 Oct;23(10):1287-93.
127. Ayer SW, McInnes AG, Thibault P, Walter JA, Doull JL, Parnell T, et al. Jadomycin, a novel 8H-benz[b]oxazolo[3,2-f]phenanthridine antibiotic from S*treptomyces venezuelae* ISP5230. Tetrahedron Letters. 1991 Oct;32(44):6301-4.
128. Han L, Yang KQ, Ramalingam E, Mosher RH, Vining LC. Cloning and characterization of polyketide synthase genes for jadomycin B biosynthesis in *Streptomyces venezuelae* ISP5230. Microbiology-Uk. 1994 Dec;140:3379-89.
129. Rubin RL, Uetrecht JP, Jones JE. Cytotoxicity of oxidative metabolites of procainamide. Journal of Pharmacology and Experimental Therapeutics. 1987;242(3):833.
130. Erikson JM, Tweedie DJ, Ducore JM, Prough RA. Cytotoxicity and DNA Damage Caused by the Azoxy Metabolites of Procarbazine in L1210 Tumor Cells. Cancer Research. 1989;49(1):127.
131. Nir I. Toxicological and Pharmacotherapeutical Aspects of Azo, Hydrazo and Azoxy Compounds. PATAI'S Chemistry of Functional Groups. 2009 2009/12/15.
132. Thirumalachar M. *Chainia*, a New Genus of the Actinomycetales. 1955;176:934-5.
133. Goodfellow M, Williams ST, Alderson G. Transfer of C*hainia* Species to the Genus S*treptomyces* with amended Description of Species. Systematic and Applied Microbiology. 1986 Jul;8(1-2):55-60.
134. Ju KS, Gao JT, Doroghazi JR, Wang KKA, Thibodeaux CJ, Li S, et al. Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes. Proceedings of the National Academy of Sciences of the United States of America. 2015 Sep;112(39):12175-80.
135. Sidda JD, Corre C. Gamma-Butyrolactone and Furan Signaling Systems in *Streptomyces*. Natural Product Biosynthesis by Microorganisms and Plants, Pt C. 2012;517:71-87.
136. Horinouchi S, Kumada Y, Beppu T. Unstable genetic determinant of A-factor biosynthesis in streptomycin-producing organisms: cloning and characterization. Journal of Bacteriology. 1984;158(2):481-7.
137. Kato JY, Funa N, Watanabe H, Ohnishi Y, Horinouchi S. Biosynthesis of gamma-butyrolactone autoregulators that switch on secondary metabolism and morphological development in *Streptomyces*. Proceedings of the National Academy of Sciences of the United States of America. 2007 Feb;104(7):2378-83.
138. Kato JY, Miyahisa I, Mashiko M, Ohnishi Y, Horinouchi S. A single target is sufficient to account for the biological effects of the A-factor receptor protein of *Streptomyces griseus*. Journal of Bacteriology. 2004 Apr;186(7):2206-11.
139. Malet N. Structure-activity relationship and biosynthesis of the methylenomycin furans [PhD]: University of Warwick; 2012.
140. Sakuda S, Tanaka S, Mizuno K, Sukcharoen O, Nihira T, Yamada Y. Biosynthetic studies on virginiae butanolide A, a butyrolactone autoregulator from *Streptomyces*. Part 2. Preparation of possible biosynthetic intermediates and conversion experiments in a cell-free system. Journal of the Chemical Society-Perkin Transactions 1. 1993 Oct(19):2309-15.
141. Shikura N, Yamamura J, Nihira T. *barS1*, a gene for biosynthesis of a gamma-butyrolactone autoregulator, a microbial signaling molecule eliciting antibiotic production in *Streptomyces* species. Journal of Bacteriology. 2002 Sep;184(18):5151-7.
142. Arakawa K. Genetic and biochemical analysis of the antibiotic biosynthetic gene clusters on the *Streptomyces* linear plasmid. Bioscience Biotechnology and Biochemistry. 2014 Feb;78(2):183-9.
143. Stratigopoulos G, Gandecha AR, Cundliffe E. Regulation of tylosin production and morphological differentiation in *Streptomyces fradiae* by TylP, a deduced gamma-butyrolactone receptor. Molecular Microbiology. 2002 Aug;45(3):735-44.
144. Corre C, Haynes SW, Malet N, Song LJ, Challis GL. A butenolide intermediate in methylenomycin furan biosynthesis is implied by incorporation of stereospecifically ^{13}C-labelled glycerols. Chemical Communications. 2010;46(23):4079-81.
145. Sidda JD, Poon V, Song LJ, Wang WS, Yang KQ, Corre C. Overproduction and identification of butyrolactones SCB1-8 in the antibiotic production superhost *Streptomyces* M1152. Organic & Biomolecular Chemistry. 2016;14(27):6390-3.
146. Biarnes-Carrera M, Lee CK, Nihira T, Breitling R, Takano E. Orthogonal Regulatory Circuits for *Escherichia coli* Based on the gamma-Butyrolactone System of *Streptomyces coelicolor*. Acs Synthetic Biology. 2018 Apr;7(4):1043-55.
147. Lougher M. Functional and Structural Insights into MmyJ, An ArsR-Like Transcriptional Repressor [PhD]: University of Warwick; 2015.
148. Zhou S. Antibiotic biosynthesis and its transcriptional regulation in *Streptomyces* bacteria: University of Warwick; 2016.
149. Styles K. Investigating Interactions Between Methylenomycin Furan Microbial Hormones And Transcriptional Repressors in *Streptomyces coelicolor* [PhD]: University of Warwick; 2016.
150. Maddocks SE, Oyston PCF. Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. Microbiology-Sgm. 2008 Dec;154:3609-23.

151. Wang Y, Rawlings M, Gibson DT, Labbe D, Bergeron H, Brousseau R, et al. Identification of a membrane protein and a truncated LysR-type regulator associated with the toluene degradation pathway in P*seudomonas putida* F1. Molecular and General Genetics. 1995 Mar;246(5):570-9.
152. Bhukya H, Jana AK, Sengupta N, Anand R. Structural and dynamics studies of the TetR family protein, CprB from *Streptomyces coelicolor* in complex with its biological operator sequence. Journal of Structural Biology. 2017 May;198(2):134-46.
153. Qiu JF, Zhuo Y, Zhu DQ, Zhou XF, Zhang LX, Bai LQ, et al. Overexpression of the ABC transporter AvtAB increases avermectin production in *Streptomyces avermitilis*. Applied Microbiology and Biotechnology. 2011 Oct;92(2):337-45.
154. Miyamoto KT, Kitani S, Komatsu M, Ikeda H, Nihira T. The autoregulator receptor homologue AvaR3 plays a regulatory role in antibiotic production, mycelial aggregation and colony development of *Streptomyces avermitilis*. Microbiology-Sgm. 2011 Aug;157:2266-75.
155. Suroto DA, Kitani S, Miyamoto KT, Sakihama Y, Arai M, Ikeda H, et al. Activation of cryptic phthoxazolin A production in *Streptomyces avermitilis* by the disruption of autoregulator-receptor homologue AvaR3. Journal of Bioscience and Bioengineering. 2017 Dec;124(6):611-7.
156. Li X, Wang J, Shi MX, Wang WS, Corre C, Yang KQ. Evidence for the formation of ScbR/ScbR2 heterodimers and identification of one of the regulatory targets in *Streptomyces coelicolor*. Applied Microbiology and Biotechnology. 2017 Jul;101(13):5333-40.
157. Orth P, Schnappinger D, Hillen W, Saenger W, Hinrichs W. Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. Nature Structural Biology. 2000 Mar;7(3):215-9.
158. Werten S, Dalm D, Palm GJ, Grimm CC, Hinrichs W. Tetracycline Repressor Allostery Does Not Depend on Divalent Metal Recognition. Biochemistry. 2014 Dec;53(50):7990-8.
159. Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM. The many faces of the helix-turn-helix domain: Transcription regulation and beyond. Fems Microbiology Reviews. 2005 Apr;29(2):231-62.
160. Hollis T. Crystallization of Protein-DNA Complexes. In: Walker JM, Doublié S, editors. Macromolecular Crystallography Protocols: Volume 1, Preparation and Crystallization of Macromolecules. Totowa, NJ: Humana Press; 2007. p. 225-37.
161. Schumacher MA, Miller MC, Grkovic S, Brown MH, Skurray RA, Brennan RG. Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR. Embo Journal. 2002 Mar;21(5):1210-8.
162. Itou H, Watanabe N, Yao M, Shirakihara Y, Tanaka I. Crystal Structures of the Multidrug Binding Repressor *Corynebacterium glutamicum* CgmR in Complex with Inducers and with an Operator. Journal of Molecular Biology. 2010 Oct;403(2):174-84.
163. Miller DJ, Zhang YM, Subramanian C, Rock CO, White SW. Structural basis for the transcriptional regulation of membrane lipid homeostasis. Nature Structural & Molecular Biology. 2010 Aug;17(8):971-U79.
164. Sawai H, Yamanaka M, Sugimoto H, Shiro Y, Aono S. Structural Basis for the Transcriptional Regulation of Heme Homeostasis in *Lactococcus lactis*. Journal of Biological Chemistry. 2012 Aug;287(36):30755-68.
165. Le TBK, Schumacher MA, Lawson DM, Brennan RG, Buttner MJ. The crystal structure of the TetR family transcriptional repressor SimR bound to DNA and the role of a flexible N-terminal extension in minor groove binding. Nucleic Acids Research. 2011 Nov;39(21):9433-47.
166. Tonthat NK, Milam SL, Chinnam N, Whitfill T, Margolin W, Schumacher MA. SlmA forms a higher-order structure on DNA that inhibits cytokinetic Z-ring formation over the nucleoid. Proceedings of the National Academy of Sciences of the United States of America. 2013 Jun;110(26):10586-91.
167. Koclega KD, Chruszcz M, Zimmerman MD, Bujacz G, Minor W. "Hot" Macromolecular Crystals. Crystal Growth & Design. 2010 Feb;10(2):580-6.
168. Yang SF, Gao ZQ, Li TT, Yang M, Zhang TY, Dong YH, et al. Structural Basis for Interaction between *Mycobacterium smegmatis* Ms6564, a TetR Family Master Regulator, and Its Target DNA. Journal of Biological Chemistry. 2013 Aug;288(33):23687-95.
169. Kim Y, Kang Y, Choe J. Crystal structure of *Pseudomonas aeruginosa* transcriptional regulator PA2196 bound to its operator DNA. Biochemical and Biophysical Research Communications. 2013 Oct;440(2):317-21.
170. Bock T, Volz C, Hering V, Scrima A, Muller R, Blankenfeldt W. The AibR-isovaleryl coenzyme A regulator and its DNA binding site - a model for the regulation of alternative *de novo* isovaleryl coenzyme A biosynthesis in *Myxococcus xanthus*. Nucleic Acids Research. 2017 Feb;45(4):2166-78.
171. Yeo HK, Park YW, Lee JY. Structural basis of operator sites recognition and effector binding in the TetR family transcription regulator FadR. Nucleic Acids Research. 2017 Apr;45(7):4244-54.
172. Bhukya H, Anand R. TetR Regulators: A Structural and Functional Perspective. Journal of the Indian Institute of Science. 2017 Jun;97(2):245-59.
173. Tovar K, Hillen W. Tet repressor binding induced curvature of *tet* operator DNA. Nucleic Acids Research. 1989 Aug;17(16):6515-22.
174. Sevvana M, Goetz C, Goeke D, Wimmer C, Berens C, Hillen W, et al. An Exclusive alpha/beta Code Directs Allostery in TetR-Peptide Complexes. Journal of Molecular Biology. 2012 Feb;416(1):46-56.

175. Kamionka A, Bogdanska-Urbaniak J, Scholz O, Hillen W. Two mutations in the tetracycline repressor change the inducer anhydrotetracycline to a corepressor. Nucleic Acids Research. 2004 Jan;32(2):842-7.
176. Schumacher MA, Zeng WJ. Structures of the nucleoid occlusion protein SlmA bound to DNA and the C-terminal domain of the cytoskeletal protein FtsZ. Proceedings of the National Academy of Sciences of the United States of America. 2016 May;113(18):4988-93.
177. Natsume R, Ohnishi Y, Senda T, Horinouchi S. Crystal structure of a gamma-butyrolactone autoregulator receptor protein in *Streptomyces coelicolor* A3(2). Journal of Molecular Biology. 2004 Feb 13;336(2):409-19.
178. Ray S, Maitra A, Biswas A, Panjikar S, Mondal J, Anand R. Functional insights into the mode of DNA and ligand binding of the TetR family regulator TylP from *Streptomyces fradiae*. Journal of Biological Chemistry. 2017 Sep;292(37):15301-11.
179. Sander T, Freyss J, von Korff M, Rufener C. Data Warrior: An Open-Source Program For Chemistry Aware Data Visualization And Analysis. Journal of Chemical Information and Modeling. 2015 Feb;55(2):460-73.
180. van Dijk M, Bonvin AMJJ. 3D-DART: a DNA structure modelling server. Nucleic Acids Research. 2009 Jul 1;37:W235-W9.
181. de Vries SJ, Bonvin AMJJ. CPORT: A Consensus Interface Predictor and Its Performance in Prediction-Driven Docking with HADDOCK. Plos One. 2011 Mar 25;6(3).
182. van Zundert GCP, Rodrigues J, Trellet M, Schmitz C, Kastritis PL, Karaca E, et al. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. Journal of Molecular Biology. 2016 Feb;428(4):720-5.
183. Marvin was used for the generation of 3D chemical structure coordinate files, Marvin 14.10.27.0, 2014, ChemAxon (http://www.chemaxon.com).
184. Plotting and non-linear regression fitting of IC_{50} and K_D analysis was performed using GraphPad Prism version 7.0a for Mac, GraphPad Software, La Jolla California USA, www.graphpad.com.
185. SnapGene software (from GSL Biotech; available at snapgene.com).
186. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. Journal of Computational Chemistry. 2009 Dec;30(16):2785-91.
187. Sanner MF. Python: A programming language for software integration and development. Journal of Molecular Graphics & Modelling. 1999 Feb;17(1):57-61.
188. Trott O, Olson AJ. Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. Journal of Computational Chemistry. 2010 Jan;31(2):455-61.
189. TL B, C E. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol. 1994;2:28-36.
190. TL B, M G. Combining evidence using p-values: application to sequence homology searches. Bioinformatics. 1998;14(1):48-54.
191. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Research. 2009 Jul;37:W202-W8.
192. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. Bioinformatics. 2011 Apr;27(7):1017-8.
193. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Molecular Biology and Evolution. 2018 Jun;35(6):1547-9.
194. The PyMOL Molecular Graphics System, Version 1.9 Schrödinger, LLC.
195. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre^2 web portal for protein modeling, prediction and analysis. Nature protocols. 2015 2015-Jun;10(6):845-58.
196. Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.695. *Department of Genome Sciences, University of Washington, Seattle*2013. p. *Distributed by the author*.
197. Grosdidier A, Zoete V, Michielin O. SwissDock, a protein-small molecule docking web service based on EADock DSS. Nucleic Acids Research. 2011 Jul;39:W270-W7.
198. Grosdidier A, Zoete V, Michielin O. Fast Docking Using the CHARMM Force Field with EADock DSS. Journal of Computational Chemistry. 2011 Jul;32(10):2149-59.
199. Invitrogen™. Champion™ pET151 Directional TOPO™ Expression Kit with BL21 Star™ (DE3) One Shot™ Chemically Competent *E. coli.* ThermoFisher.com: ThermoFisher Scientific; 2010 [cited 2018 09/09/18].
200. Stanton BC, Nielsen AAK, Tamsir A, Clancy K, Peterson T, Voigt CA. Genomic mining of prokaryotic repressors for orthogonal logic gates. Nature Chemical Biology. 2014 Feb;10(2):99-105.
201. Ho NAT, Dawes SS, Crowe AM, Casabon I, Gao C, Kendall SL, et al. The Structure of the Transcriptional Repressor KstR in Complex with CoA Thioester Cholesterol Metabolites Sheds Light on the Regulation of Cholesterol Catabolism in *Mycobacterium tuberculosis*. Journal of Biological Chemistry. 2016 Apr;291(14):7256-66.
202. Hill AV. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. J Physiol (Lond). 1910 1910;40:4-7.

203. Heuer H, Krsek M, Baker P, Smalla K, Wellington EMH. Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients. Applied and Environmental Microbiology. 1997 Aug;63(8):3233-41.
204. Onaka H, Horinouchi S. DNA-binding activity of the A-factor receptor protein and its recognition DNA sequences. Molecular Microbiology. 1997 Jun;24(5):991-1000.
205. Majka J, Speck C. Analysis of Protein–DNA Interactions Using Surface Plasmon Resonance. In: Seitz H, editor. Analytics of Protein–DNA Interactions. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007. p. 13-36.
206. Engohang-Ndong J, Baillat D, Aumercier M, Bellefontaine F, Besra GS, Locht C, et al. EthR, a repressor of the TetR/CamR family implicated in ethionamide resistance in mycobacteria, octamerizes cooperatively on its operator. Molecular Microbiology. 2004 Jan;51(1):175-88.
207. Carette X, Blondiaux N, Willery E, Hoos S, Lecat-Guillet N, Lens Z, et al. Structural activation of the transcriptional repressor EthR from *Mycobacterium tuberculosis* by single amino acid change mimicking natural and synthetic ligands. Nucleic Acids Research. 2012 Apr;40(7):3018-30.
208. Craus C, Willand N, Villemagne B, Flipo M, Willery E, Carette X, et al. Unconventional surface plasmon resonance signals reveal quantitative inhibition of transcriptional repressor EthR by synthetic ligands. Analytical Biochemistry. 2014 May;452:54-66.
209. Lin J, Cagliero C, Guo BQ, Barton YW, Maurel MC, Payot S, et al. Bile salts modulate expression of the CmeABC multidrug efflux pump in *Campylobacter jejuni*. Journal of Bacteriology. 2005 Nov;187(21):7417-24.
210. Muhl D, Jessberger N, Hasselt K, Jardin C, Sticht H, Burkovski A. DNA binding by *Corynebacterium glutamicum* TetR-type transcription regulator AmtR. Bmc Molecular Biology. 2009 Jul;10:13.
211. Baucheron S, Coste F, Canepa S, Maurel MC, Giraud E, Culard F, et al. Binding of the RamR Repressor to Wild-Type and Mutated Promoters of the *ramA* Gene Involved in Efflux-Mediated Multidrug Resistance in *Salmonella enterica* Serovar Typhimurium. Antimicrobial Agents and Chemotherapy. 2012 Feb;56(2):942-8.
212. Anand S, Singh V, Singh AK, Mittal M, Datt M, Subramani B, et al. Equilibrium binding and kinetic characterization of putative tetracycline repressor family transcription regulator Fad35R from *Mycobacterium tuberculosis*. Febs Journal. 2012 Sep;279(17):3214-28.
213. Petridis M, Vickers C, Robson J, McKenzie JL, Bereza M, Sharrock A, et al. Structure and Function of AmtR in *Mycobacterium smegmatis*: Implications for Post-Transcriptional Regulation of Urea Metabolism through a Small Antisense RNA. Journal of Molecular Biology. 2016 Oct;428(21):4315-29.
214. Kim Y, Kim BS, Park YJ, Choi WC, Hwang J, Kang BS, et al. Crystal Structure of SmcR, a Quorum-sensing Master Regulator of *Vibrio vulnificus*, Provides Insight into Its Regulation of Transcription. Journal of Biological Chemistry. 2010 Apr;285(18):14020-30.
215. Krell T, Teran W, Mayorga OL, Rivas G, Jimenez M, Daniels C, et al. Optimization of the palindromic order of the TtgR operator enhances binding cooperativity. Journal of Molecular Biology. 2007 Jun;369(5):1188-99.
216. Liang JL, Nie Y, Wang MX, Xiong GM, Wang YP, Maser E, et al. Regulation of alkane degradation pathway by a TetR family repressor via an autoregulation positive feedback mechanism in a Gram-positive *Dietzia* bacterium. Molecular Microbiology. 2016 Jan;99(2):338-59.
217. Ballok AE, Bahl CD, Dolben EL, Lindsay AK, Laurent JDS, Hogan DA, et al. Epoxide-Mediated CifR Repression of *cif* Gene Expression Utilizes Two Binding Sites in *Pseudomonas aeruginosa*. Journal of Bacteriology. 2012 Oct;194(19):5315-24.
218. Fujihashi M, Nakatani T, Hirooka K, Matsuoka H, Fujita Y, Miki K. Structural characterization of a ligand-bound form of *Bacillus subtilis* FadR involved in the regulation of fatty acid degradation. Proteins-Structure Function and Bioinformatics. 2014 Jul;82(7):1301-10.
219. Su CC, Rutherford DJ, Yu EW. Characterization of the multidrug efflux regulator AcrR from *Escherichia coli*. Biochemical and Biophysical Research Communications. 2007 Sep;361(1):85-90.
220. Hoffmann KM, Williams D, Shafer WA, Brennan RG. Characterization of the multiple transferable resistance repressor, MtrR, from *Neisseria gonorrhoeae*. Journal of Bacteriology. 2005 Jul;187(14):5008-12.
221. Garcia-Nafria J, Baumgart M, Turkenburg JP, Wilkinson AJ, Bott M, Wilson KS. Crystal and solution studies reveal that the transcriptional regulator AcnR of *Corynebacterium glutamicum* is regulated by citrate-Mg^{2+} binding to a non-canonical Pocket. Journal of Biological Chemistry. 2013 May;288(22):15800-12.
222. Ahn SK, Tahlan K, Yu Z, Nodwell J. Investigation of transcription repression and small-molecule responsiveness by TetR-like transcription factors using a heterologous *Escherichia coli*-based assay. Journal of Bacteriology. 2007 Sep;189(18):6655-64.
223. Dam TK, Roy R, Pagé D, Brewer CF. Negative Cooperativity Associated with Binding of Multivalent Carbohydrates to Lectins. Thermodynamic Analysis of the “Multivalency Effect”. Biochemistry. 2002 2002/01/01;41(4):1351-8.
224. Dam TK, Gabius H-J, André S, Kaltner H, Lensch M, Brewer CF. Galectins Bind to the Multivalent Glycoprotein Asialofetuin with Enhanced Affinities and a Gradient of Decreasing Binding Constants. Biochemistry. 2005 2005/09/01;44(37):12564-71.

225. Balasubramanian S, Sorokulova IB, Vodyanoy VJ, Simonian AL. Lytic phage as a specific and selective probe for detection of Staphylococcus aureus - A surface plasmon resonance spectroscopic study. Biosensors & Bioelectronics. 2007 Jan;22(6):948-55.
226. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. Quantifying similarity between motifs. Genome Biol. 2007;8(2):R24.
227. Bunet R, Mendes MV, Rouhier N, Pang X, Hotel L, Leblond P, et al. Regulation of the synthesis of the angucyclinone antibiotic alpomycin in Streptomyces ambofaciens by the autoregulator receptor AlpZ and its specific ligand. Journal of Bacteriology. 2008 May;190(9):3293-305.
228. Bunet R, Song LJ, Mendes MV, Corre C, Hotel L, Rouhier N, et al. Characterization and Manipulation of the Pathway-Specific Late Regulator AlpW Reveals Streptomyces ambofaciens as a New Producer of Kinamycins. Journal of Bacteriology. 2011 Mar;193(5):1142-53.
229. Kurniawan YN, Kitani S, Iida A, Maeda A, Nijeholt JLA, Lee YJ, et al. Regulation of production of the blue pigment indigoidine by the pseudo gamma-butyrolactone receptor FarR2 in Streptomyces lavendulae FRI-5. Journal of Bioscience and Bioengineering. 2016 Apr;121(4):372-9.
230. Folcher M, Gaillard H, Nguyen LT, Nguyen KT, Lacroix P, Bamas-Jacques N, et al. Pleiotropic functions of a Streptomyces pristinaespiralis autoregulator receptor in development, antibiotic biosynthesis, and expression of a superoxide dismutase. Journal of Biological Chemistry. 2001 Nov;276(47):44297-306.
231. Mast Y, Guezguez J, Handel F, Schinko E. A Complex Signaling Cascade Governs Pristinamycin Biosynthesis in Streptomyces pristinaespiralis. Applied and Environmental Microbiology. 2015 Oct;81(19):6621-36.
232. Liu WS, Zhang QL, Guo J, Chen Z, Li JL, Wen Y. Increasing Avermectin Production in *Streptomyces avermitilis* by Manipulating the Expression of a Novel TetR-Family Regulator and Its Target Gene Product. Applied and Environmental Microbiology. 2015 Aug;81(15):5157-73.
233. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 2004 Mar;32(5):1792-7.
234. Salehi-Najafabadi Z, Barreiro C, Martinez-Castro M, Solera E, Martin JF. Characterisation of a gamma-butyrolactone receptor of Streptomyces tacrolimicus: effect on sporulation and tacrolimus biosynthesis. Applied Microbiology and Biotechnology. 2011 Dec;92(5):971-84.
235. Sugiyama M, Onaka H, Nakagawa T, Horinouchi S. Site-directed mutagenesis of the A-factor receptor protein: Val-41 important for DNA-binding and Trp-119 important for ligand-binding. Gene. 1998 Nov;222(1):133-44.
236. Onaka H, Sugiyama M, Horinouchi S. A mutation at proline-115 in the A-factor receptor protein of *Streptomyces griseus* abolishes DNA-binding ability but not ligand-binding ability. Journal of Bacteriology. 1997 Apr;179(8):2748-52.
237. Davis JB, Bailey JD, Sello JK. Biomimetic Synthesis of a New Class of Bacterial Signaling Molecules. Organic Letters. 2009 2009/07/16;11(14):2984-7.
238. Morin JB, Adams KL, Sello JK. Replication of biosynthetic reactions enables efficient synthesis of A-factor, a γ-butyrolactone autoinducer from *Streptomyces griseus*. Organic & Biomolecular Chemistry. 2012;10(8):1517-20.
239. Keay BA. Synthesis of multi-substituted furan rings: the role of silicon. Chemical Society Reviews. 1999 Jul;28(4):209-15.
240. Fischer R, Fiserova M. One-step synthesis of furan-2,5-dicarboxylic acid from furan-2-carboxylic acid using carbon dioxide. Arkivoc. 2013:405-12.
241. Bowyer JE, de los Santos ELC, Styles KM, Fullwood A, Corre C, Bates DG. Modeling the architecture of the regulatory system controlling methylenomycin production in *Streptomyces coelicolor*. Journal of Biological Engineering. 2017 Oct;11:12.
242. Biarnes-Carrera M, Breitling R, Takano E. Butyrolactone signalling circuits for synthetic biology. Current Opinion in Chemical Biology. 2015 2015/10/01/;28:91-8.
243. Weber W, Schoenmakers R, Spielmann M, El-Baba MD, Folcher M, Keller B, et al. Streptomyces-derived quorum-sensing systems engineered for adjustable transgene expression in mammalian cells and mice. Nucleic acids research. 2003;31(14):e71-e.
244. Folcher M, Xie M, Spinnler A, Fussenegger M. Synthetic mammalian trigger-controlled bipartite transcription factors. Nucleic acids research. 2013;41(13):e134-e.
245. Fox BG, Blommel PG. Autoinduction of protein expression. Current protocols in protein science. 2009;Chapter 5:Unit-5.23.
246. Studier FW. Stable Expression Clones and Auto-Induction for Protein Production in E. coli. In: Chen YW, editor. Structural Genomics: General Applications. Totowa, NJ: Humana Press; 2014. p. 17-32.
247. Briand L, Marcion G, Kriznik A, Heydel JM, Artur Y, Garrido C, et al. A self-inducible heterologous protein expression system in *Escherichia coli*. Scientific Reports. 2016 Sep;6:11.
248. Nocadello S, Swennen EF. The new pLAI (lux regulon based auto-inducible) expression system for recombinant protein production in *Escherichia coli*. Microbial Cell Factories. 2012 Jan;11:10.
249. Bulheller BM, Hirst JD. DichroCalc-circular and linear dichroism online. Bioinformatics. 2009 Feb;25(4):539-40.
250. Geertz M, Maerkl SJ. Experimental strategies for studying transcription factor–DNA binding specificities. Briefings in Functional Genomics. 2010;9(5-6):362-73.

251. Yang YH, Joo HS, Lee K, Liou KK, Lee HC, Sohng JK, et al. Novel method for detection of butanolides in *Streptomyces coelicolor* culture broth, using a His-tagged receptor (ScbR) and mass spectrometry. Applied and Environmental Microbiology. 2005 Sep;71(9):5050-5.
252. Joo HS, Yang YH, Lee CS, Kim JH, Kim BG. Fragmentation study on butanolides with tandem mass spectrometry and its application for the screening of S&R-captured quorum sensing molecules in *Streptomyces coelicolor* A3(2). Rapid Communications in Mass Spectrometry. 2007;21(5):764-70.
253. Ueda K, Beppu T. Antibiotics in microbial coculture. Journal of Antibiotics. 2017 Apr;70(4):361-5.
254. Klose DP, Wallace BA, Janes RW. 2Struc: the secondary structure server. Bioinformatics. 2010 Oct;26(20):2624-5.